



## **Microeconomic methods applied in relation to food consumption, health and obesity prevention**

Hansen, Aslak Hedemann

*Publication date:*  
2011

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Hansen, A. H. (2011). *Microeconomic methods applied in relation to food consumption, health and obesity prevention*. Institute of Food and Resource Economics, University of Copenhagen.

# Microeconomic Methods Applied in Relation to Food Consumption, Health and Obesity Prevention

---

Aslak Hedemann Hansen

PhD thesis submitted as fulfillment of the requirements for the degree of Doctor of Philosophy

## Acknowledgements

This research was undertaken as an activity within the Danish Obesity Research Centre (DanORC), financed by the Danish Strategic Research Council.

I gratefully acknowledge the financial support of the Danish Strategic Research Council.

## Abstract

---

Obesity rates are rising worldwide. Consequently, vast amounts of resources are invested in an attempt to curtail this development. Policy evaluation can be a cost effective measure in so far as it can help to direct resources towards the most effective solutions. However, this of course implicitly assumes that the policy evaluation is reliable.

The use of the evaluation results can also be limited if they do not reflect population heterogeneity and reduction in health disparity is a policy goal in so far as policy implementation which is directed at specific target groups is hampered by the restricted information.

Hence, the reliability and value of the evaluation results is constrained by the data collection process and the statistical methods employed. This thesis is concerned with methodological issues regarding the policy evaluation of measures aimed at modifying food consumption and obesity. The methodologies considered aim to reveal population heterogeneity which is important if policies are to be targeted with maximum effect and if health disparities are a concern.

The thesis consists of six papers four of which are concerned with food consumption, obesity and policy evaluation.

The first paper introduces the topic of the thesis. The objective of this paper is to discuss why obesity is a problem, why it exists, policies to solve the problem and policy evaluation. The paper reviews some policy instruments currently in use. The paper is especially concerned with policy evaluation: methodological issues and issues related to population heterogeneity and inequality in health. The paper concludes that one of the main problems of obesity is its negative effects on productivity and the fiscal economy, which is increasingly important in an ageing European population.

The paper also finds that policy evaluation could be improved if methodological and data restrictions are removed.

Regarding evaluations of real policies, the paper finds that policy evaluations are not always reliable especially simple comparisons of pre-policy and post-policy surveys, which are highly

prevalent. This provides evidence for the assertion that it is important that policy evaluation does not merely measure the change in the sample of a quantity of interest, but that it also accounts for changes in the composition of the distribution of socio-demographic variables, such as education and income, whilst accounting for the development in prices if appropriate. Policy evaluations that do not account for such factors can give rise to misleading results. The paper suggests, in this case, to use a matching method such as the one applied in paper 4, which is attractive in that it solves the aforementioned problem and also reflects population heterogeneity in the measure of interest. It is, therefore, a useful tool for measuring developments in specific target groups such as individuals who don't eat much fruit and vegetables, or eat too much saturated fatty acids.

Regarding evaluations of hypothetical policy experiments based on historical data, the paper identifies a number of restrictions on most of the models applied in the literature and notes that some of these restrictions can be lifted if appropriate data are available such as data consisting of individual household level time series including food consumption information, i.e. dietary information and price, and socio-demographic status. The analysis in Paper 5 lifted one of these restrictions, namely a restriction on heterogeneity.

Papers 2 and 3 are methodological papers that deal with issues related to Paper 4.

Paper 2 entitled, *A data problem in a popular dataset - Impact and solution*, deals with a problem in the data, GfK Consumer Tracking data, used in the thesis and delivered by the **market research** company GfK Denmark. The data problem is only an issue in paper 4. The implication of the data problem is that the values of prices in the data have to be estimated. An analysis of a naïve estimator formerly used is undertaken to assess the seriousness of the problem. Finally, a methodology based on the Expectation Maximisation (EM) algorithm is developed which gives much better results than the naïve estimator. The solution developed in this paper is used to cleanse the fruit and vegetables data used in paper 4.

Paper 3 deals with the *estimation of multi-dimensional discrete stochastic variables*. The accuracy of the methodology used in paper 4 depends on how well the covariate distribution of the multi-dimensional discrete stochastic variables is estimated. Previous applications used a maximum likelihood estimator, however, it is well known that when the number of observations is small relative to the number of parameters that are to be estimated, it can be an advantage in finite sample applications to use a method that smoothes parameter estimates using a kernel function. Least square cross-validation is an example of such a smoothing approach which leads to differing asymptotic behaviour of the smoothing parameters depending on whether each variable is

uniformly distributed or not. Also, smoothening parameters, which correspond to uniformly distributed variables, have positive probability of not converging to their optimal value when using least square cross-validation. This paper shows why this is the case and derives the asymptotic probability distribution of the smoothening parameter. Furthermore, a criterion function is suggested that solves this problem. The fact that the smoothening parameters now asymptotically converge to the optimal values effectively means that the associated estimator shrinks to the smallest variance unbiased frequency estimator and lends support to the usability of smoothening methods in discrete variable applications.

Paper 4 examines the *determinants of development in inequality in fruit and vegetable consumption in Denmark*. Inequality in consumption of fruit and vegetables (FAV) in Denmark has increased in a period of extensive information campaigns, which aimed to increase the consumption of FAV. The paper attempts to identify the reasons for this development. Quantile regression is used on a rich data set from the market research institute GfK Denmark for the analysis of the development in FAV consumption. The determinants of the development in FAV consumption are investigated by use of a matching methodology that enables comparisons of data samples with different covariate distributions. The study shows a negative relationship between the intensity of consumption and price sensitivity. Increased inequality in consumption over time can partly be attributed to this negative relationship coupled with increases in prices, but to a greater extent so that uneducated groups with a low consumption and low income groups are falling behind the more educated and financially well off segment of the population. The findings may be utilised to target specific sub-populations when designing policies aimed at increasing fruit and vegetable consumption. These findings highlight the advantages of quantitative methodologies that compare developments in samples and adjust for differences in covariate distributions.

Paper 5 examines *consumption effects of a tax on saturated fat in foods*. The purpose of this paper was to explain how one can forecast the effect of a proposed tax on saturated fat (SFAs) on the demand for butter and margarine. The tax is supposed to come into effect in mid 2011 in Denmark. The tax was created to affect the consumption of SFAs and it is therefore of special interest to know whether the population is price sensitive. Earlier studies simply focused on the mean price response, but this is unsatisfactory because most populations have a significant variation in their responses. Instead, the individual household's price response is estimated here by unlocking more information

on the expected response to the tax reform. The main purpose was to examine whether a tax on SFAs can improve the Danes' diet, while taking the possible population heterogeneity into account. The possibility of answering this question was obtained by taking advantage of a detailed panel dataset, and using a theoretical framework that excels by estimating household specific prices, and deals with missing prices by taking into account the preferences of households. Also, the estimation of the demand system at the household level puts less restrictive assumptions on the estimates than previous studies. A great heterogeneity within the population price responsiveness underlines the importance of the adopted approach and indicates that consumers should be treated accordingly when considering the development of new food taxes. The analysis reveals that a likely outcome of the tax for the majority of the population is a decrease in the purchase of hard margarines and butter products, and that these will, to some extent, be substituted with healthier alternatives, such as spreadable margarine, minarine and vegetable oils. It is therefore expected that the tax reform will help to improve the Danes' fat intake profile, and consequently the health of the Danish population. Furthermore, the analysis reveals that the proposed SFAs tax is superior to a VAT on the considered food items in terms of modifying the consumers' food consumption behaviour in a desirable way.

Paper 6 examines a hypothesis of an existence of *non-linearities in consumers' demand response to price changes* using Bayesian estimation techniques. The analysis extends the empirical demand analysis by searching for potential non-linearities in the response to price changes. In general, the empirical analysis suggests that the demand for drinking milk responds significantly to price changes. But, the analysis also shows some asymmetry in this response to price changes. In particular, price decreases tend to trigger larger demand responses than price increases.

# Resumé

---

Fedme er stigende på verdensplan. Derfor investeres enorme ressourcer i et forsøg på at begrænse denne udvikling. Evaluering af politikker kan være en omkostningseffektiv foranstaltning for så vidt som det kan bidrage til at determinere de mest effektive løsninger. Men dette forudsætter naturligvis, at evalueringen er pålidelig. Brugen af evalueringsresultaterne kan også være begrænsede, hvis de ikke afspejler befolkningens heterogenitet og reducere i social ulighed i sundhed er en politisk målsætning, for så vidt som gennemførelsen af politikkerne, der er rettet mod specifikke målgrupper hæmmes af den begrænsede information. Derfor er pålideligheden og værdien af evalueringsresultaterne begrænset af kvaliteten af de indsamlede data og statistiske metoder. Denne afhandling beskæftiger sig med metodologiske spørgsmål vedrørende evaluering af foranstaltninger med henblik på at ændre forbruget af fødevarer og fedme. De anvendte metoder har til formål at afsløre heterogenitet i befolkningen, som er vigtig, hvis politikkerne skal være målrettet med maksimal effekt, og hvis ulighed i sundhed er et problem.

Afhandlingen består af seks artikler hvoraf fire vedrører fødevarerforbrug, fedme og evaluering af politikker.

Den første artikel introducerer emnet for afhandlingen. Formålet med dette oplæg er at diskutere, hvorfor overvægt er et problem, hvorfor det eksisterer, politikker til at løse problemet og evaluering af politikker. Artiklen gennemgår nogle politiske instrumenter i brug. Artiklen fokuserer især på evaluering: metodiske spørgsmål og spørgsmål i forbindelse med heterogenitet i befolkningen og ulighed i sundhed. Artiklen konkluderer, at en af de største problemer ved fedme er dens negative virkning på produktiviteten og den finanspolitiske økonomi, som bliver stadig vigtigere i en aldrende europæisk befolkning.

Artiklen konstaterer endvidere, at evaluering af politikkerne kunne blive forbedret, hvis begrænsninger på metoder og data fjernes.

Med hensyn til evaluering af reelle politikker, konstaterer artiklen, at evalueringer ikke altid er pålidelige, især simple sammenligninger af stikprøver før og efter implementering af politikken, som er meget udbredt. Dette giver støtte til påstanden om, at det er vigtigt, at evaluering af en givet politik ikke alene måler ændringer i stikprøven af en variabel af interesse, såsom indtaget af frugt og grønt, men at den også tager højde for ændringer i sammensætningen af fordelingen af socio-

demografiske variabler, såsom uddannelse og indkomst, og for udviklingen i priserne, hvis nødvendigt. Politikevalueringer, der ikke tager højde for sådanne faktorer kan give anledning til misvisende resultater. Artiklen foreslår, i dette tilfælde, at bruge en ”matching” metode som den, der anvendes i artikel 4, som er attraktiv, da den løser det nævnte problem og også afspejler befolkningens heterogenitet i variabelen af interesse. Den er derfor et nyttigt værktøj til at måle udviklingen i bestemte målgrupper, såsom personer, der ikke spiser meget frugt og grønt, eller spiser for meget mættet fedt.

Med hensyn til evalueringer af hypotetiske politikeksperimenter baseret på historiske data, peger artiklen på en række restriktioner på de fleste af de anvendte modeller i litteraturen, og bemærker, at nogle af disse restriktioner kan ophæves, hvis der foreligger passende data såsom data, der består af tidsserier af individuelle husstande herunder oplysninger om fødevareforbrug, pris og socio-demografisk status. Analysen i artikel 5 løfter en af disse begrænsninger, nemlig en begrænsning på heterogenitet i populationen.

Artikel 2 og 3 er metodiske artikler, der beskæftiger sig med spørgsmål relateret til artikel 4. Artikel 2 med titlen, *Et data problem i et populært datasæt - Virkning og løsning*, beskæftiger sig med et problem i de data, GfK Consumer Tracking data, der anvendes i afhandlingen og leveret af markedsanalysefirmaet GfK Danmark. Dataproblemet er kun et problem i artikel 4. Følgerne af dataproblemet er, at det er nødvendigt at estimere værdierne af priserne i data. En analyse af en naiv estimator der tidligere blev anvendt, er foretaget for at vurdere omfanget af problemet. Endelig er en metode baseret på en ”Expectation Maximisation” (EM) algoritme udviklet, som giver langt bedre resultater end den naive estimator. Løsningen udviklet i denne artikel er brugt til at rense de frugt og grønt data, der anvendes i artikel 4.

Artikel 3 omhandler *Estimation af en multi-dimensional diskret stokastisk variabel*. Nøjagtigheden af den metode, der anvendes i artikel 4 afhænger af, hvor godt fordelingen af den socio-økonomiske variabel, som er en multi-dimensional diskret stokastisk variabel, er estimeret. Tidligere applikationer brugte en maximum likelihood estimator, dog er det velkendt, at når antallet af observationer er lille i forhold til antallet af parametre, der skal estimeres, kan det være en fordel at anvende en metode der udglatte parameter estimererne via en vægtfunktion. Mindste kvadraters krydsvalidering er et eksempel på en sådan udglatnings fremgangsmåde, der fører til forskellig asymptotisk opførsel for udglatningsparametre afhængigt af, om variabelen er ensartet fordelt eller



ej. Derudover gælder, at udglatningsparametre, der svarer til ensartet fordelte variable, har positiv sandsynlighed for ikke at konvergere til deres optimale værdi, når mindste kvadraters krydsvalidering anvendes. Artikel 3 viser, hvorfor dette er tilfældet, og udleder den asymptotiske sandsynlighedsfordeling for udglatningsparametrene. Desuden foreslås en kriteriefunktion, der løser dette problem. Den omstændighed, at udglatningsparametre nu asymptotisk konvergerer til de optimale værdier, betyder, at den tilknyttede estimator krymper til den mindste varians bias-fri frekvens estimator og giver støtte til anvendeligheden af udglatningsmetoder i diskrete variable applikationer.

Artikel 4 undersøger determinanter for udviklingen i uligheden i frugt og grønt forbruget i Danmark. Ulighed i forbruget af frugt og grøntsager (FOG) i Danmark er steget i en periode med omfattende informationskampagner, som havde til formål at øge forbruget af FOG. Artiklen forsøger at finde årsagerne til denne udvikling. Fraktil regression anvendes på et rigt datasæt fra markedsanalysefirmaet GfK Danmark til analysering af udviklingen i FOG forbruget. Determinanterne for udviklingen i FOG forbruget undersøges ved anvendelse af en "matching" metode, der muliggør sammenligninger af stikprøver med forskellige socio-økonomiske fordelinger. Undersøgelsen viser en negativ sammenhæng mellem intensiteten af forbruget og prisfølsomhed. Øget ulighed i forbruget over tid kan delvis tilskrives denne negative sammenhæng kombineret med stigninger i priserne, men i højere grad, at uuddannede grupper med et lavt forbrug og lavindkomstgrupper sakker bagud i forhold til den veluddannede og økonomisk velstillede segment af befolkningen. Resultaterne kan bruges til at målrette politikker mod specifikke delpopulationer der sigter mod at øge indtaget af frugt og grøntsager. Disse resultater fremhæver fordelene ved kvantitative metoder, som sammenligner udviklingen i stikprøver og justerer for forskelle i fordelingen af den socio-økonomiske variabel.

Artikel 5 undersøger *Forbrugs virkninger af en afgift på mættet fedt i fødevarer*. Formålet med denne artikel var at forklare, hvordan man kan forudsige effekten af en foreslået afgift på mættet fedt på efterspørgslen efter smør og margarine. Afgiften skal træde i kraft i midten af 2011 i Danmark. Afgiften blev oprettet for at påvirke forbruget af mættet fedt og det er derfor af særlig interesse at vide, om befolkningen er prisfølsomme. Tidligere undersøgelser har fokuseret på den gennemsnitlige pris respons, men det er utilfredsstillende, fordi de fleste populationer har en betydelig variation i deres respons. I stedet er den enkelte husholdnings pris reaktion estimeret, og herved frigøres mere information om den forventede reaktion på skattereformen. Hovedformålet var

at undersøge, om en afgift på mættet fedt kan forbedre danskernes kost, samtidig med at en eventuel heterogenitet i pris responsen i befolkningen tages i betragtning. Muligheden for at besvare dette spørgsmål blev opnået ved at drage fordel af et detaljeret panel datasæt, og ved hjælp af en teoretisk ramme, der udmærker sig ved at estimere husstand specifikke priser, og estimere manglende priser ved at tage hensyn til husstandes individuelle præferencer. Estimering af efterspørgselssystemet på husstandsniveau lægger mindre restriktive antagelser på estimationerne end tidligere undersøgelser. En stor uensartethed i befolkningens prisfølsomhed understreger betydningen af den valgte fremgangsmåde og indikerer, at forbrugere skal behandles i overensstemmelse hermed, når man overvejer at udvikle nye fødevarerafgifter. Analysen viser, at et sandsynligt resultat af afgiften for størstedelen af befolkningen er et fald i køb af hårde margariner og smørprodukter, og at disse i et vist omfang kan erstattes med sundere alternativer, såsom smørbar margarine, minarine og vegetabiliske olier. Det forventes derfor, at skattereformen vil bidrage til at forbedre danskernes fedtindtagsprofil, og dermed sundheden af den danske befolkning. Derudover viser analysen, at den foreslåede mættet fedt skat er bedre end en moms baseret skat på de betragtede fødevarer til at ændre forbrugernes fødevareforbrugsadfærd i en ønskelig retning.

Artikel 6 undersøger en hypotese om eksistens af *ikke-lineariteter i forbrugernes efterspørgselsreaktion på prisændringer* ved hjælp af Bayesian estimeringsmetoder. Analysen udvider de traditionelle empiriske efterspørgselsanalyser ved at søge efter potentielle ikke-lineariteter i efterspørgselsreaktionen ved prisændringer. Generelt indikerer den empiriske analyse, at forbrugernes efterspørgsel efter konsummælk reagerer betydeligt ved prisændringer. Men analysen viser også asymmetri i dette respons på prisændringer. I særdeleshed er der en tendens til at prisfald udløser større efterspørgselsrespons end prisstigninger.

# Contents

---

Abstract  
Resumé  
Contents

Paper 1: Obesity Causation, Policy Solution and Evaluation

Paper 2: A data problem in a popular dataset - Impact and solution

Paper 3: Estimation of multi-dimensional discrete stochastic variables

Paper 4: Consumption effects of a tax on saturated fat in foods

Paper 5: Determinants of development in inequality in fruit and vegetable consumption in Denmark

Paper 6: Non-linearities in consumers' demand response to price changes - the case of milk demand

# Paper 1: Obesity Causation, Policy Solution and Evaluation

---

## **Abstract**

The objective of this paper is to discuss why obesity is a problem, why it exists, policies to solve the problem and policy evaluation. The dominant economic view is the neoclassical one that technological developments have changed the human condition with implied behavioural changes that have disturbed individuals' energy balance equation. The paper reviews some policy instruments in use. The paper is especially concerned about policy evaluation: methodological issues and issues related to population heterogeneity and inequality in health. Finally, I draw some conclusions on how future policy evaluation could be improved.

## **Background**

The prevalence of being overweight and obese is increasing at a rapid pace. According to WHO (2006) figures, approximately 1.6 billion adults were overweight, and at least 400 million adults were obese globally in 2005, whilst by 2015, approximately 2.3 billion adults will be overweight and more than 700 million will be obese. The EU conforms to this trend, according to the European Commission (Platform Charter, 2005), European Union citizens are experiencing, "a sustained, acute EU-wide increase in overweight and obesity" (European Commission, 2007). The International Obesity Taskforce (IASO/IOTF, 2010) estimates that in the European Union 27 member states, approximately 60% of adults and over 20% of school-age children are overweight or obese. This equates to around 260 million adults and over 12 million children who are either overweight or obese.

Improving the diet quality of the European population, as well as other measures designed to contain or reverse current obesity trends, has consequently become a priority and is now on the political agenda in the EU.

## ***Organisation of the paper***

The paper is organised as follows. First, I discuss why obesity is a problem. Why should we as economists care about the obesity epidemic? Secondly, I consider explanations for obesity including the neoclassical view (rational agents) and behavioural economic arguments, such as self-control problems. I then review some policy instruments currently in use. I then review methods used for policy evaluation and studies using these methodologies. Finally, conclusions connected to policy evaluation are drawn based on the review.

## Obesity: a worldwide problem

Obesity is a problem for several reasons. Seen from the perspective of the individual, obesity affects quality of life and health, since overweight and obesity in adults, are linked to increased risk of heart disease, type 2 diabetes, high blood pressure, certain cancers, and other chronic conditions (WHO Global Health Risks Report, 2004; U.S. Department of Health and Human Services, 2003).

The disability attributable to obesity and its consequences in 2004 was calculated at over 36 million disability-adjusted life years (DALYs), due primarily to ischaemic heart disease and type 2 diabetes (WHO Global Health Risks Report, 2004). In 2004, overweight and obesity was estimated to account for 2.8 million deaths, whilst the combined total with physical inactivity was 6.0 million. This total accounts for 11 per cent of responsible deaths globally, very close to the top risk factor for death, high blood pressure at 13 per cent (WHO Global Health Risks Report, 2004).

Besides the value of lost life, obese individuals incur other costs, including lost wages, gasoline costs, and, when applicable, life insurance and health care costs (Dor et al., 2010).

Obesity can lead to lower income possibly due to a perception on behalf of the employer that the obese are less productive, because of more sick days etc., (Dor et al., 2010; Bhattacharya, et al., 2009). This perception of lower productivity is not ungrounded according to a study by Østbye *et al.* (2007) who examined the number and type of employees' compensation claims from Duke University and found a clear linear relationship between body mass index (BMI) and the rate of claims. Employees with a BMI over 40 filed twice as many workers' compensation claims as employees with a recommended-weight, whilst they had more than 12 times as many lost work days. While many studies support the conjecture that obesity can lead to lower income, it is less clear whether lower income will also lead to obesity. However, a recent study showed that earning minimum wages significantly increases the risk of becoming obese (DaeHwan et al., 2010). For a review of the economic explanations for the observed correlation between BMI and socioeconomic status see Dam, Jensen & Kærgård (2008).

From a welfare economic perspective, the disadvantages of being obese are reason enough to intervene if a significant proportion of obese individuals, or individuals at risk of becoming obese, would prefer not to be obese. Behavioral economics is implicitly present in this reasoning as it is used to justify paternalism: people are irrational, or incapable of judging the tradeoffs involved in their decision making, hence they do not make decisions that serve their own interests, which justifies intervention.

Furthermore, the obesity epidemic puts a considerable strain on the fiscal economy, e.g. 10-15 percent of the lost years of life in Europe and a significant corresponding loss in tax revenue can be attributed to poor nutrition. This is worsened by obesity's contribution of as much as 2-6 percent of the total direct health care costs in several European countries, which is mainly financed by government expenditure (WHO, 2003).

On top of this, under most circumstances/systems, obese individuals will impose economic burdens, externalities, on normal weight individuals, i.e. when the financing of the health care service an individual receives is independent of his weight. This argument is the neoclassical case for intervention and is not only valid in Europe as the U.S government pays a significant portion of the direct medical costs through publicly funded programs such as Medicare and Medicaid (Dor et al. 2010). The neoclassical argument for intervention has been criticised for overlooking potential offsetting fiscal effects of increased obesity, such as reduced fiscal spending on public pension annuity programmes. Taxpayers may actually save due to increased obesity, because obese individuals have higher mortality which implies a higher risk

of early termination of the payment stream of their annuity contract (Lakdawalla and Philipson, 2006) and lifetime medical costs of obese individuals might also be lower than for normal weight individuals, again because of excess mortality (van Baal PHM, Polder JJ, de Wit GA, Hoogenveen RT, Feenstra TL, *et al.*, 2008).

Obese individuals might not impose negative externalities through their contribution of taxes relative to their health care and pension expenses, and the prevention of obesity might not be a cure for increasing health expenditure, nevertheless prevention of obesity might still contribute positively to the fiscal economy through its labour increasing effect. This is relevant because of the combination of ageing populations and a contracting domestic labour force which will have drastic consequences for Europe (Project Europe 2030 on the Challenge of Demography, 2010). The growth in productivity would allow for a revised allocation of resources that could help fill the increasing gap between pension receivers and contributors. However, the exact status of how obesity prevention affects the fiscal economy is unresolved, because of the lack of studies that include all of its effects including the direct benefits such as reduced health care expenditures and indirect benefits, e.g. productivity boosts, etc. More importantly, studies are needed that take into account the added value of obesity prevention which results from increases in labour market participation in Europe through its labour productivity increasing effect. Labour market participation in Europe is expected to rise in the future as it is, “a sine qua non” (Project Europe 2030 on the Challenge of Demography, 2010).

In addition to the individual and the government, employers bear some of the costs of obesity. Employers and employees share the burden of many costs including short-term disability, absenteeism and productivity losses. Employers directly pick up the costs for many of these expenditures. However, employees indirectly share part of this burden through lower wages (Dor et al., 2010). There are also costs associated with the infrastructure changes that societies must construct to handle the obesity epidemic, i.e. larger sized hospital beds, operating tables and wheel chairs are needed; enlarged seats at sports-grounds, and modifications to transport safety standards (IASO/IOTF, 2010).

### ***Obesity's inequality creating effect***

Governments are interested in inequality in the population because of the stability of society and moral concerns. Lakdawalla and Philipson (2006) argued that taxpayers may actually save due to increased obesity, although their argumentation does not account for indirect costs such as productivity decreases. The authors speculate that the neoclassical externality argument is not the true rationale for an intervention, but a convenient substitute for other implicit rationales such as the inequality creating effect of the obesity epidemic.

The obesity epidemic has incited a non-uniformly distributed obesity increase over socio-demographic groups because today there is a clear social gradient in obesity. This is problematic because low income and less well-educated individuals have worse health than their wealthy counterparts independent of weight. Note that a uniformly distributed obesity increase over socio-demographic groups implies increased health inequality within socio-demographic groups and a non-uniformly distributed obesity increase, in addition, increases health inequality across socio-demographic groups. Hence, the inequality creating effect of the obesity epidemic is a rationale for intervention.

## **Understanding the causation of obesity**

The use of the energy balance equation for understanding the causation of obesity is highly prevalent. Its basis on the

thermodynamic laws provides exact quantitative relationships between energy intake, energy expenditure and deposition of energy which is translated into body-weight changes. However, the model cannot separate the initiating and driving forces of the energy imbalance, which may also be an active storage of fat in adipose tissue. These and various other limitations of the energy balance model warrant cautiousness when using the model in studies of obesity causation (Sørensen, 2009). The dominant view is that it is changes in the human condition and associated changes in behaviour which have increased the probability that an individual will enter a state of positive energy balance, thereby increasing the risk of becoming obese. For example, the WHO (2000) ascribe the obesity epidemic to increased sedentary lifestyles and high-fat, energy-dense diets. This is despite the existence of evidence of a predisposition to gain weight for certain individuals, e.g. as a result of postnatal environmental influences (Sørensen, 2009), so that they consequently have a higher probability of becoming obese than others, and despite the limitations of the energy balance model.

### ***Obesity driven by changes in economic incentives***

During the development of the obesity epidemic, different technological shocks have occurred that have changed the human condition and altered economic incentives in the process. For instance, a less energy demanding and more sedentary lifestyle has been the result of automated transportation and technology induced shifts in labour demand from the primary and secondary sector to the service sector, whilst the remaining labour in the primary and secondary sectors now expend less energy as a result of new technology. See also Lakdawalla and Philipson (2002), Cutler *et al.* (2003) and Finkelstein *et al.* (2005) for further arguments that link technological progress with individuals' economic incentives for choosing an obesity creating lifestyle.

On the other side of the energy balance equation, increased energy consumption has been ascribed to *intensive advertising of unhealthy foods* (The PolMark Project, 2010; Nestlé, 2002; French *et al.*, 2001), *increasing portion sizes and energy density* (Martin, 2005; Berg *et al.*, 2009; Seidell, 2009; Vermeer *et al.*, 2009; Kral & Rolls, 2004) *in both pre-packaged food and in restaurants and fast food outlets* (Wansink, 2004; French *et al.*, 2001), *increasing work hours for women* (Scholdere, 2007), *increasing consumption of convenience, fast food, unhealthy snack foods and sugar-sweetened beverages* (Walker *et al.*, 2008; Malik *et al.*, 2006; French *et al.*, 2000), *reduced food prices through industrial innovation for mass preparation of foods containing added fat and sugar* (Lakdawalla & Philipson, 2002; Powell & Chaloupka, 2009) and *changes in the relative cost of healthy compared to unhealthy food* (Cutler *et al.*, 2003; Finkelstein *et al.*, 2005) *including reduced time costs in food preparation*.

## Self-control problems – a behavioural economics explanation of the rise in obesity

Insights into food consumption decisions can be derived from behavioural economics theories and psychological studies. Especially what has been labelled time inconsistent choices or dynamically inconsistent preferences has been the subject of intense studies. The concept of time inconsistent choices is easily understood within the context of dieting. It is easy for an obese individual to decide to go on a weight-loss diet; recognizing the long-term advantages of dietary restraint and how they outweigh the short-lived advantages of caloric indulgence, it is a benefit tradeoff. However, the food that was so easy to forgo as an abstraction is not so easy to forgo when it is right in front of you, i.e. the tradeoff is no longer acceptable when the time arrives. Indeed, the decision to abstain from eating forbidden (fattening) foods, followed by the collapse of this promise when the forbidden food is actually encountered, is a perfect example of dynamically inconsistent preferences (Herman & Polivy, 2003). Economists compare intertemporal tradeoffs by use of a discount factor (see Modigliani, 1975). Let  $\beta_t$  be the benefit of “caloric indulgence” at time  $t$  and  $U_t$  be the utility of decreased weight at time  $t+1$ .<sup>1</sup> Now, since the obese individual is willing to go on a diet at time  $t$  he has a personal discount factor at time  $t$ ,  $\delta_t$ , that satisfies:  $\beta_t \geq \delta_t U_t$ , but since he considers it unacceptable to start a diet when the time arrives, he has a personal discount factor at time  $0$ ,  $\delta_0$ , that satisfies:  $\beta_0 < \delta_0 U_0$ .<sup>2</sup> Note that the two inequalities have either the implication  $\delta_t > \delta_0$  or  $\delta_t < \delta_0$ . Both of these implications are incompatible with the assumptions in neoclassical theory of intertemporal choice. Assume that  $\delta_t > \delta_0$  and  $\delta_t < \delta_0$ , this discount structure sets up a conflict between today’s preferences, and the preferences that will be held in the future. For example, from today’s point of view, the discount rate between two far-off periods,  $t$  and  $t+1$ , is the long-term low discount rate. However, from the time  $t$  point of view, the discount rate between  $t$  and  $t+1$  is the short-term high discount rate. Although this type of preference change reflects dynamically inconsistent preferences, they are reflected in many familiar experiences. For example, today I might aspire to start a resolute exercise regime (or diet) tomorrow, but when tomorrow arrives, my preference is to procrastinate and postpone any sacrifices to another day. To accommodate this type of phenomenon, it is necessary to relax the assumption of a constant discounting factor (exponential discounting in continuous time) in the neoclassical theory of intertemporal choice. Hence, the introduction of hyperbolic discount functions (in continuous time and quasi-hyperbolic discount functions in discrete time; Laibson, 1997) in behavioural economics, which allows the discount factor to decay over time. Models with a decaying interpersonal discount factor are substantiated by biological evidence on the structure of the brain, which uses rather diverse neural systems to manage immediately presented rewards (the limbic system) from those that are used to manage rewards at different times in the more distant future (the prefrontal cortex), McClure *et al.* (2004).

The theory and empirical evidence of the existence of humans with self-control problems, i.e. individuals with dynamically inconsistent preferences, can help explain the rise in obesity in so far as changes in the human condition, such as technological developments, have aggravated a tendency among individuals to procrastinate and put off dieting and exercising and to choose immediate gratification over long term health. As previously stated, this incapability of

<sup>1</sup> Note it is assumed that eating has implications for the weight at the next point in time  $t+1$ , i.e.  $\beta_t = \beta_{t+1}$ .

<sup>2</sup> Note it is assumed that the tradeoffs are identical except that they occur at different points in time, i.e.  $U_t = U_{t+1}$ ,  $\beta_t = \beta_{t+1}$ ;



judging intertemporal tradeoffs is a potential justification for government intervention regarding individuals with self-control problems. Nordström & Thunström (2009) describe this excellently, “If they do not commit themselves to behaving fully rationally, their immediate actions impose externalities on their future selves.”

In the previous paragraph, increasing portion sizes was mentioned as one of the possible reasons for the rise in obesity. As an example of how consumers might end up with inappropriate portion sizes and how the food industry’s marketing mix can affect obesity, Vermeer, Steenhuis & Seidell (2009) found that overweight fast food customers are more likely to choose small soft drink sizes and less likely to choose large food portion sizes when the price was related linearly to quantity. Thus, quantity discounts might be the straw that broke the camel's back for overweight consumers with self-control problems. For further examples of how self-control problems and other behavioural economic phenomena, such as habitual behaviour and irrational choices, can help explain the observed rise in obesity, see Dam, Jensen & Kærgård (2008).

To mitigate such self-control problems, Laibson (1997) allows the decision maker to foresee these conflicts and uses a commitment technology to partially limit the options available in the future.

## **Food items associated with obesity and disease**

There are several studies that try to identify food items that are associated with obesity and health. In the following, some of the evidence regarding saturated fatty acids and fruit and vegetables is described.

First, evidence links obesity with a diet containing a relatively high intake of energy-dense foods and saturated fatty acids (SFAs). Therefore, policies that aim at limiting SFAs, or decreasing the SFAs to unsaturated fatty acids ratio, are considered likely to improve obesity prevalence rates (WHO, 2003). This evidence was translated into recommendations and guidelines in the WHO Global Strategy on Diet, Physical Activity and Health (WHO, 2003b, 2004), which was adopted in May 2004. The WHO recommends, “moving from saturated animal-based fats to unsaturated vegetable-oil based fats” (WHO, 2003b), to “limit energy intake from total fats and to shift fat consumption away from saturated fats to unsaturated fats” (WHO, 2004). Recently, Hasselbalch *et al.* (2010) found that the intake of vegetable oil was inversely related to waist change and the European Food Safety Authority (EFSA, 2009; p.2) put forward a recommendation concerning the intake of saturated fatty acids in the EU population, “The Panel recommends that SFA intake should be as low as possible within the context of a nutritionally adequate diet.” However, recent meta-analysis based studies provide evidence to support the hypothesis that what is more important is what the saturated fatty acids are substituted with, rather than merely reducing the SFA intake

(Mozaffarian et al., 2010), i.e. the risk of coronary heart disease (CHD) can be reduced by the substitution of SFAs with poly unsaturated fatty acids (PUFAs). A similar conclusion was reached by an expert panel review of the evidence of SFA as a risk factor (Astrup et al., 2011). Astrup et al. (2011) found consistent evidence that the risk of CHD is reduced when SFAs are replaced by PUFAs, but no clear benefit of substituting carbohydrates with SFAs has been shown, although there might be a benefit if the carbohydrate is unrefined and has a low glycemic index. Astrup et al. (2011) further argue that the effect of particular foods on CHD cannot be predicted solely by their content of total SFAs because individual SFAs may have different cardiovascular effects, whilst major SFA food sources contain other constituents that could influence CHD risk. Astrup *et al.* (2011) finally conclude that research is needed to compare specific foods with appropriate alternatives.

The learning outcome of this literature review may be that it might not be a good idea to address SFAs in foods *per se*. Rather it would be better to address foods containing SFAs where better alternatives exist, e.g. a SFA tax on butter could be beneficial, because a healthier alternative, rapeseed oil, exists. An “appropriate alternative” should satisfy multiple requirements. The chief one is of course that it should be healthier relative to the alternatives, but it should also be possible to introduce it into a delicious diet considering the cultural tradition, whilst it should also be relatively inexpensive. The latter requirements can be relaxed by information campaigns and subsidies.

Secondly, a WHO/FAO<sup>3</sup> expert consultation report on diet, nutrition and the prevention of chronic diseases recommends an intake of a minimum of 400 g of fruits and vegetables (FAV) per day for the prevention of chronic diseases such as heart disease, cancer, diabetes and obesity WHO/FAO(2003). The report states that there is convincing evidence that fruit and vegetables decrease the risk of obesity. Eating a variety of vegetables and fruit ensures an adequate intake of dietary fibres and can help displace foods which are high in saturated fats and sugar. There is convincing evidence that low-glycemic index foods, such as FAV and foods rich in whole grain, are associated with a reduced risk of type 2 diabetes and cardiovascular disease. There is evidence of a link between gut microbial metabolism and key factors associated with insulin resistance (Nilsson et al., 2008, Nilsson et al., 2010). Studies have shown that dietary fibre may increase and maintain satiety and postpone the onset of hunger (Nilsson et al., 2008, Bosch et al., 2009) and the inclusion of fermentable fibre in canine diets may contribute to the prevention or mitigation of obesity through its effects on satiety (Bosch et al., 2009). Finally, low FAV intake is present in the top 8 leading risk factor causes of death in middle and high income countries (WHO, 2004).

## Policies

The energy balance equation is most often implicitly present in the policy debate concerning the instruments that might be useful in counteracting obesity. Most policy recommendations seek to minimise the risk of an individual ending up

---

<sup>3</sup> World Health Organization/Food and Agricultural Organization.

in a state of positive energy balance through education to emphasise the importance of exercise and diet, or through providing incentives to spend more energy and/or modifying the food consumption behaviour favourably. Many policies try to fix or reverse what is believed to be the cause of the rise in obesity, i.e. the changes in the human condition caused by technological developments. For instance, if technological developments have resulted in humans spending less energy on travelling then a policy goal could be to increase the amount of active travelling.

#### *Instruments that seek to increase energy expenditure by physical activity*

A possible instrument that could increase energy expenditure might be to allow the doctors (general practitioners) of obese individuals to prescribe 'Exercise on prescription' (EoP). A study of five similar EoP programmes concludes that EoP can contribute to improvements in physical activity levels and health-related quality of life for physically inactive patients with lifestyle diseases, or those who are at increased risk of developing lifestyle diseases (Sørensen *et al.*, 2010). Another example is the public-private partnerships (PPP) "Cycle to Work" in Iceland, which is a partnership between The National Olympic and Sports Association of Iceland (head organiser), the Public Health Institute of Iceland, cycle clubs and The Icelandic National Broadcasting Service. The objective of the partnership is to promote active transport (walking and cycling) to work (European Commission, 2008). Fifteen EU Member States have either partly or fully implemented policy actions to promote active travel (Jakab, 2010). Also, the PPP "Girls on the Move" established in April 2005, aims to improve the physical activity levels of adolescent girls and young women in Scotland. Furthermore, the European Heart Network commitment has a project executed by the Danish Heart Foundation that addresses aerobic fitness by educating and physically testing children in schools in Denmark (EU platform on diet, physical activity and health, 2010). According to this data, 25 % of children have very low fitness levels implying an increased risk of developing cardiovascular disease and the metabolic syndrome<sup>4</sup>. The project financed by the Ministry of Health in Denmark was inexpensive. Project participants are now working to get all Danish children tested every year, and are considering the possibility of scaling the project up across Europe. Twenty-four out of twenty-seven EU Member States have either partly or fully implemented guidelines regarding physical activity (Jakab, 2010).

However, most suggested policy instruments seek to modify the food consumption environment in a way that favours food consumption behaviour that prevents obesity.

### ***Public-private partnerships (PPP)***

The EU commission has recommended public-private partnerships (PPP) as a platform whereby members can make commitments to contribute to the pursuit of healthy nutrition, physical activity and the fight against obesity (European Commission, 2008). Many member states have PPP which are focused on healthy life-styles, especially the prevention of overweight and obesity through a balanced diet and physical activity. For instance, the Czech Republic's "Keep it Balanced" and "Action Plan on Counteracting Overweight and Obesity;" Germany's initiative for a healthy diet and more physical activity; Italy's, "GAIN HEALTH"; The national Covenant on overweight and the Rotterdam Covenant

---

<sup>4</sup> In prospective cohort studies, higher levels of cardiorespiratory fitness have rather consistently protected against the development of diabetes and cardiovascular disease, conditions that are frequently linked with the metabolic syndrome (Laaksonen *et al.*, 2002).

On Nutrition and Physical Activity in the Netherlands; the Polish Platform for Action on Diet, Physical Activity and Health; Slovakia's National Program on Preventing Obesity and the Portuguese Platform Against Obesity (European Commission, 2008).

## *Advertising*

It is generally believed that advertising can influence food choices; arguably the level of resources allocated to marketing is in itself proof that it works, otherwise the industry would not spend such vast amounts of resources on marketing (The PolMark Project, 2010; Nestlé, 2002). In particular, marketing which is directed towards children has raised concern (The PolMark Project, 2010). In Quebec for example, a ban on advertising to children under thirteen has been in effect since 1980. The ban constitutes a natural experiment for policy evaluation, as no ban was enacted in the neighbouring province of Ontario. Baylis & Dhar (2007) use difference-in-difference (DID) and matching methods to estimate the effect of the ban. The results revealed a drop in the number of fast food meals purchased ranging from 11 to 22 million or 8.9 to 23 billion calories if measured in energy.

## *Information campaign*

Public information campaigns are in the group of healthy eating interventions which apply commercial marketing technologies to influence the eating behaviour of target segments in order to improve their personal welfare and that of society. Thus, a campaign aims to empower consumers with information which is hoped will affect actual behaviour. Wakefield *et al.* (2010) review the outcomes of mass media campaigns in the context of various health-risk behaviours and conclude that mass media campaigns can produce positive changes or prevent negative changes in health-related behaviours across large populations. However, the evidence is not completely unambiguous as Rickertsen & von Cramon-Taubadel (2003) find it difficult to find any effects of health information on food demand in the Scottish population by including a health-information index in a food-demand model. Information campaigns might not always be effective because they are frequently competing with other factors, such as powerful social norms, pervasive product marketing and behaviours driven by habit or addiction. Wakefield *et al.*, (2010) propose investment in longer term better-funded campaigns to achieve adequate population exposure to media messages.

An example of a successful information campaign constructed within a PPP is the Short TV programs constructed within the Food Industry Initiative in the framework of the Charter to promote healthy diet and physical activity in programmes and advertising in France 2010-2011 (Chapalain, 2010). The PPP is a partnership between the Ministry of Health, the Ministry of Culture and the private sector, such as TV companies and the food industry. The aim of the short programs constructed within the PPP is to promote a healthy diet and physical activity by targeting the young population in television programs and advertising. The health messages were validated and controlled by scientists and experts in social marketing. As part of the partnership, television companies granted special pricing for state advertising, e.g. up to 60% tariff reductions for the National Institute for Health Prevention and Education, and all French television companies participated. The market research company TNS' evaluation of season1 (2010) was very positive, especially concerning stated behaviour improvements. The objective of the evaluation was to measure the efficiency of the short TV programs, with an emphasis on the impact of the short programs concerning visibility and

memorisation; to verify that the messages were clearly understood i.e. that children and parents had understood the link between good health and a good diet; and to judge the effectiveness of the campaign in creating incentives for children and adults to adopt healthy behaviour regarding food and physical activity. 500 women and 350 children aged 6 to 12 years participated in the survey. 87% of women agreed that the campaign “Makes me want to change my family food habits” and 86% of children agreed that “it gives me desire to participate in physical activity.”

The U.S. Department of Health (2004) stresses the importance of communicating nutrition messages, including the National Cancer Institute's 5-A-Day for Better Health Program to increase FAV consumption in order to address the obesity epidemic.

Many countries have adopted government supported campaigns aimed at increasing FAV consumption: “5 a day” is such an example and it has been adopted in various English speaking countries, notably the United States (Havas *et al.*, 1995) as well as England, Wales and Scotland. In Denmark, a similar campaign is the “6 a day” campaign, which was agreed upon by central nutrition educators in September 1998 (Ministry of Food, Agriculture and Fisheries, 2001). It is a public-private partnership with representatives from government agencies, non-governmental health organisations, consumer organisations and industrial sectors, including the FAV industry and the meat industry<sup>5</sup>. The primary goal of the campaign is to increase the consumption of FAV among Danish consumers (European Commission, 2008). Many other countries have implemented a 5-a-day campaign or similar campaigns including France, Spain, Poland and Canada.

The European Commission (2005) is interested in identifying measures that could contribute towards improving the attractiveness, availability, accessibility and affordability of fruit and vegetables. A prominent example of a policy at the supranational level is the reform within the Common Agricultural Policy of the Common Market Organisation (CMO) for fruit and vegetables, which is aimed at promoting FAV consumption within specific settings, such as schools. Part of the reform will promote children's consumption of FAV through proposals to allow surplus production to be distributed to educational institutions and children's holiday centres. Additionally, a school fruit scheme offering greater affordability of FAV to encourage consumption is part of the reform of the CMO for fruit and vegetables (European Commission, 2007). Policies aimed at increasing FAV consumption are by far the most common policy. It is symptomatic that the provision of subsidised school fruit schemes is implemented by the majority of EU countries, 26 out of 27 EU Member States either partly implemented or fully implemented a school fruit scheme (Jakab, 2010)

### **Labelling**

Some EU PPP involve nutrition labelling, e.g. in Romania the partnership “Common nutrition labelling scheme;” in Spain the partnership entitled, “Food labelling, food composition & advertising;” the Danish Wholegrain logo and in Sweden, “Keyhole labelling” and “Keyhole at restaurants.” In general, the goal of these partnerships is to help people make informed dietary choices and diet and lifestyle choices for good health. One of the Spanish partnership's main objectives is to include nutritional information and labelling on products. In Romania, the partnership is working for a

---

<sup>5</sup> The Danish Meat Association contribute to 6-a-day with knowledge about meat as part of a meal, culinary cooking, recipe development, etc. and by providing network contacts such as dieticians, teachers of home economics, food writers etc.

National Nutrition Labelling Scheme in working groups based on requirements of Directive 2000/13/EC on labelling, presentation and advertising of foodstuffs and CIAA guideline daily amounts (European Commission, 2008). The Swedish Keyhole labelling partnership was established in 1989 with the aim of enabling consumers to make healthier choices when buying foods. The National Food Administration has developed criteria for the keyhole labelling (revised in 2006) in dialogue with retailing and the food industry. Initially, only the fibre and fat content was included in the criteria, but in the 2006 revision, several other nutrients were included. The use of the label is voluntary, but the criteria are regulated (LIVSFS 2005:9). Also, in Sweden the “Keyhole at restaurants” partnership was established in 2007 with the aim of facilitating consumers’ ability to choose restaurants which serve healthier foods. A restaurant can apply for a diploma in order to be able to market the restaurant as a Keyhole Restaurant if all the employees have taken a course organised by the National Food Administration and the restaurant fulfils certain criteria regarding the foods served, e.g. energy, fat and salt content (European Commission, 2008). The purpose of the Danish wholegrain partnership is to encourage Danes to consume more whole grains, the aim being at least 75 grams a day. The PPP involves Health organisations, authorities and commercial partners. The most notable outcome of the partnership is the development of a new wholegrain logo, introduced in January 2009, which is on a range of different whole grain foods thereby helping the Danish population to choose healthy whole grain products (European Commission, 2008). Finally, the nutrition labelling programme, “*Pick the Tick*” of the National Heart Foundation of New Zealand allows food manufacturers, whose products meet defined nutritional criteria, to display the *Pick the Tick* logo on food labels. The tick is used by 59% of consumers. *Pick the Tick* is used by the food industry as a tool for marketing food products and it has provided an incentive to improve the nutritional value of foods. The tick on approved products not only acts as a ‘nutrition signpost’ for consumers, but has also significantly influenced reformulation (Young & Swinburn, 2002). Legislation about labelling energy content in foods has been implemented in all EU member states, although 10 member states have only partly implemented this policy action (Jakab, 2010).

### **Product reformulation**

WHO (1990, 2003, 2004) linked diet and disease, including increased risk of obesity from intake of energy-dense foods, and recommended governments to initiate discussions with the food industry and consumers to ensure that companies limit the levels of saturated fats, trans-fatty acids, free sugars and salt in existing products and develop new products that are low in fat, free sugars, and salt and that have a better nutritional value. WHO (2008) states that, “reformulation of food is considered as one of the key options for achieving dietary goals” and the Strategy for Europe on Nutrition Overweight and Obesity related health issues, adopted in May 2007, identifies reformulation as an important measure (European Commission, 2007). Reformulation is usually defined as the modification of the nutritional composition of foods, primarily aimed at the reduction of fats, sugars and salt.

National policies and PPP can shape reformulation directly, e.g. through a ban on trans fatty acids, or indirectly, e.g. by shifts in consumer preferences.

As already mentioned, the labelling *Pick the Tick* in New Zealand has had an effect on reformulation, and labelling regulation could have far reaching consequences. For instance, the introduction of the health claim “zero trans fat” in labelling has triggered changes all along the processed food chain, including investments in new processing technologies and the development of soy and canola crop varieties with different oil characteristics (Golan, Krissoff &

Kuchler, 2007). The mandated disclosure of trans fatty acid (TFAs) content on food labels in Canada has resulted in a decline in the intake of TFAs (L'Abbe *et al.*, 2009). Despite the apparent benefits of replacing TFAs with other fatty acids in food products, concern has been raised that this change might increase the intake of SFAs. However, data show that the TFAs in these foods can be substituted with a mixture of SFAs, MUFAs, and PUMAs. The nutritional benefit of substituting TFAs with this mixture of fatty acids for popular foods is even greater than the benefit of a one-to-one substitution of TFAs with SFAs (Stender, Astrup & Dyerberg, 2009).

All but one EU member state have either partly or fully implemented food based dietary guidelines (Jakab, 2010). The introduction of new dietary guidelines could have impacts on reformulation. For instance, the US 2005 dietary guidelines offered quantitative recommendations for the consumption of whole-grains and Mancino, Kuchler & Leibtag (2008) hypothesised that the changed recommendations were responsible for subsequent increases in retail sales and consumption of whole-grain food products. The authors found that the new dietary guidelines and related media interest increased availability and sales of whole-grain foods and a large impact on consumption occurred through reformulation, induced by product competition. Thus, policies that shift consumer preferences or in general affect consumer demand are possible drivers of reformulation. For instance, Santarossa & Mainland (2003) argue that introducing a tax to diminish consumers' fat intake will cut households' expenditure on the considered foods to such an extent that food producers will voluntarily reformulate food products towards more healthy alternatives. However, voluntary reformulation might be slow as a review of health-related reporting of twenty-five of the world's largest food companies revealed. Among the commitments and practice regarding reformulation and portion size, eight reported action on TFAs, five reported action on sugar, four reported action on fat, two reported action on portion size and only one company was taking action on all four measures (Lang, Rayner & Kaelin, 2006). An alternative to voluntary reformulation is legislation, for instance Denmark became the first country to ban the sale of foods containing industrially produced TFAs in both restaurants and grocery stores from 1 June 2003 (L'Abbe *et al.*, 2009). Also, New York City, Philadelphia, and California have banned the use of TFAs in foods prepared in restaurants, and several other states are considering similar laws (Niederdeppe & Frosch, 2009).

### ***Policies aimed at mitigating self-control problems***

While no policy program (for obvious reasons) explicitly states that one of its objectives is to mitigate self-control problems, it is clear that a number of them implicitly do so by introducing elements of commitment for participants including rewards. For instance, the PPP "Girls on the Move" was designed to tackle the barriers which discourage girls and young women from participating in physical activity. The program focuses on participation and leadership and offers participants the opportunity to take part in a wide range of physical activities along with opportunities to gain a nationally recognised qualification in leisure and recreation leadership (European Commission, 2008). People might be less likely to procrastinate and put off exercising if other individuals depend on their participation and they themselves derive utility from exercising/participating with others. Social relationships were very important to the girls participating in Girls on the Move – 82 % indicated that they were taking part with friends, and by the end of the project 79 % indicated that they had made new friends while participating in the project (Taylor, 2008).



## *Policies targeting disadvantaged groups*

In Denmark, a reduction in inequality in health is a goal of the current government and was also a goal of the previous left-wing government (Government, 2002), whilst it is also a goal of specific campaigns such as the 6-a-day campaign (Ministry of Food, Agriculture and Fisheries, 2001). The UK five-a-day campaign also includes a reduction in health inequalities as one of its policy objectives (Capacci, 2010). In the US, one of the main goals in Healthy People 2010 was to “eliminate health disparities” (US Department of Health and Human Services, 2000). The Warsaw Declaration signed by the Ministers of Health from 12 countries on 26th November 2007, aims to reduce health disparities across Europe (EU platform on diet, physical activity and health, 2010). There are quite large disparities in inequality in different dimensions of health in Europe. For example, the Eurothine study estimated that 26% and 50% of obesity prevalence for men and women in Europe respectively could be ascribed to inequalities in educational status (Roskam & Kunst, 2007). A notable example of a PPP program targeting disadvantaged groups is “Girls on the Move” in Scotland. The partnership was initiated partly in response to the gender gap in participation in physical activity among children and adolescents, as approximately 50 % more boys than girls in the age range 14-15 years were reaching the recommended amount of physical exercise in 1998 (Taylor, 2008). “Girls on the Move” was designed to raise the physical activity level of girls and young women in Scotland and to engage harder to reach groups, including girls with disabilities or mental health issues, girls displaying “at risk” behaviour, young mothers and girls from ethnic minorities and disadvantaged and deprived communities in positive physical activities (European Commission, 2008).

## **Programme and policy evaluation**

As was clearly evident from the review in the previous section, many countries spend a considerable amount of resources on nutrition and health programmes and policies.

This paragraph concerns two types of programme evaluation problems: (i) assessing the impacts of interventions of various policy experiments for which there is no prior experience, including their impacts on population health and (ii) the impact of the intervention on measures related to the health of the treated<sup>6</sup>.

The two types of problems are considered separately in the next section. The first problem requires the use of historical data to forecast the consequences of new policies. The second problem is often solved in a simple manner, e.g. by before and after comparison, but is also treated in the academic literature where the problem has been under scrutiny in the statistical treatment effect literature.

Policy evaluation is concerned with assessing the effects of a specific treatment. Ideally, such evaluation should be done by randomised controlled trials (RCT). However, without a randomised controlled trial, as is often the case, treatment assignment, i.e. participation in a programme, is non-random, the samples are drawn from different populations and failure to correct for the resulting baseline differences can lead to misleading results (Jones, 2008; Gordon *et al.*, 2006).

---

<sup>6</sup> Heckman (2008) also discusses a third programme evaluation type: forecasting the impacts of intervention implemented in one environment in other environments, but this type is not discussed in this paper.



When dealing with data which does not come from a RCT, there are different econometric models that make necessary assumptions to identify treatment effects. For example, Grootendorst (2007) review the use of instrumental variables (IV) estimation of treatment effects in the applied health sciences, and if the data can be considered a natural experiment, then a difference-in-difference (DID) regression model is ideal, as discussed by Schreyögg & Grabka (2010). However, the necessary conditions for applying IV and DID methods are rarely satisfied (see Baser, 2009), which is perhaps why the statistical technique of matching has become an accepted tool for the evaluation of treatments in studies where data is not sampled via a RCT. As is often the case with developments in econometrics, matching techniques originate from the statistics literature, but the econometrics literature has contributed with theoretical and particularly empirical work (e.g. Ho *et al.*, 2007; Cobb-Clark & Crossley, 2003). In brief, matching mimics a randomised experiment *ex post*.

Most evaluations do not take into account how the intervention effect is distributed over the population. However, a cost benefit analysis should ideally include these distributional effects because benefits should be measured in utility and the marginal utility of a given effect is different for different individuals, e.g. the benefits of an increase in FAV consumption for an individual depends on his pre-intervention intake. The evaluated effect of an intervention on social welfare can, therefore, be markedly altered if it does not take distributional effects into account. Also, disparities in health are a serious concern for governments all over the world, which makes distributional effects of an intervention even more important. Some papers which consider this problem include the evaluation in paper 5, which is based on a demand system, and evaluations based on quantile regression - e.g. Gustavsen & Rickertsen (2006, 2008), Stewardson (2003), and paper 4 - which are all discussed in detail below.

### *Applying matching techniques in the evaluation of policy intervention*

Policy evaluation often consists of a simple comparison of pre-policy and post-policy surveys. However, this can often be misleading as the following examples show.

First, the evaluation of the US 5-a-day program through two surveys (1991 and 1997) delivered a statistically significant improvement in consumption of FAV. However, the adjusted analysis revealed that the improvement was most likely attributable to changes in the composition of distribution of socio-demographic variables in the two samples (Stables *et al.*, 2002).

Secondly, the first major community-based project for coronary heart disease (CHD) prevention is the Finnish North Karelia Program. It was launched in 1972 in response to the record high mortality of CVD in the region. The Program included specific dietary changes, such as an increase in the consumption of fruits and vegetables, and shifts in consumption from saturated to unsaturated fat. Comprehensive activities included media campaigns, actions via the health service and others, schools and multi-sector collaboration including NGOs, the food industry, agriculture etc. The activities were further implemented across the whole country. The gap in age-adjusted CHD mortality rates between males aged 35-64 years in North Karelia and the whole of Finland has narrowed considerably from 1969 to 2001, which suggests that the measures of the Karelia Project have been significantly better at reducing CHD than the nationwide

activities (page 6 of Puska, 2002)<sup>7</sup>. However, North Karelia was a low socio-economic area with a scarcity of medical resources and with many socioeconomic problems in the 1970s (Puska, 2002). Hence, if the population of North Karelia has reduced its socioeconomic problems such that there is no significant disparity between North Karelia and the rest of Finland in 2001, these changes, rather than the Karelia Project, may account for the reduction of the gap in CHD. However, it seems likely that the Karelia Project has been beneficial, but an analysis that adjusts for possible changes in the sample distribution of socioeconomic variables would be necessary to reliably evaluate the effectiveness of the Karelia Project, i.e. a counterfactual estimate to assess the treatment effect (the effectiveness) of the Karelia Project.

Thirdly, Capacci & Mazzocchi (2010) recently provided an ex-post assessment of the UK 5-a-day information campaign and found that all impacts are larger than those observed by simply comparing pre-policy and post-policy intakes.

Finally, the study in chapter 4 similarly finds that the adjusted (behavioural) change is larger than the sample change at all quantiles, which suggests that the impact of the Danish 6-a-day campaign is larger than what can be observed by simply comparing pre-policy and post-policy intakes.

The examples provide evidence to support the assertion that it is important that policy evaluations, such as the evaluation of the 6-a-day campaign, do not merely measure the change in the sample of a quantity of interest, but also account for changes in the composition of distribution of socio-demographic variables, such as education and income, and account for the development in prices, if appropriate. Policy evaluations that do not account for such factors can give rise to misleading results. The examples, therefore, show the benefits of matching techniques in the evaluation of policy intervention.

### *Forecasting the consequences of new policies using historical data*

Taxes and subsidies are often suggested by economists as ways of modifying the consumption behaviour of consumers, so that the food they consume is closer to being optimal individually and socially in terms of the risk of obesity.

This has been actualised by the recently proposed fat tax in Denmark; as a means to reduce the population's saturated fatty acids (SFAs) intake, which is based on the content of SFAs in foods and is under development. This fat tax is special since it not only taxes certain products, but also introduces a new element; the taxation is based on a specific nutrient contained in food products, SFAs. The tax is scheduled to take effect in July 2011 (The Danish Ministry of Taxation, 2011 a, 2011 b, 2009 b). However, at the moment, no government has explicitly introduced taxes of significant size which aim at modifying diet behaviour, although many governments manage individual differential value added tax (VAT) regimes for food products. For example, in Denmark, an elected tax increase of 25% on ice cream, chocolate and candy, as well as a tax on soft drinks differentiated with respect to sugar, came into effect in January 2010 (The Danish Ministry of Taxation, 2009 a). Romania is another EU member state which is planning to tax fatty and sugary food. However, the proposed tax is very different from the Danish tax as it is directed towards the fast food sector (Charter, 2010).

---

<sup>7</sup> [http://www.who.int/chp/media/en/north\\_karelia\\_successful\\_ncd\\_prevention.pdf](http://www.who.int/chp/media/en/north_karelia_successful_ncd_prevention.pdf)

Only one EU Member State has fully implemented policy measures that affect food prices (Jakab, 2010).

Given the lack of existing policy interventions of significant size which have been in effect for a sufficient amount of time, evaluations or forecasts have been based on simulation using records of estimated price responsiveness of different foods.

When analysing healthy eating behaviour, specific target groups such as individuals who do not eat much, are of more interest than groups, which have an average level of consumption. This explains why quantile regression has become increasingly popular when analysing eating behaviour. Gustavsen & Rickertsen (2006, 2008) use a censored linear quantile model, which takes the zero purchase in the considered survey periods into account. In their latter paper, Gustavsen, Jolliffe & Rickertsen (2008) simulate the effect on purchases of ice cream of an increase in the VAT for less healthy foods and the removal of the VAT for healthy foods. They found that households with high purchase intensity cut their annual per capita purchases by 1.8 kilograms, which is equivalent to an annual drop of more than half a kilogram of body weight. Most studies of food consumption behaviour, however, are based on modelling mean responses.

The demand behaviour, with regard to fats, and the assessment of the effects of SFAs-taxation has been the subject of a number studies in the literature. For example, Yen & Chern (1992) and Koc *et al.* (2001) estimate price and expenditure elasticities for a number of fat products. Huang & Lin (2000) estimate food demand and nutrient elasticities on the basis of household survey data. Nichele (2003) estimates a set of behavioural elasticities (with respect to expenditure, price and health information) for a range of food products and links the responses in the consumption of these products to the intake of various lipids, including SFAs, using a set of nutrient coefficients for the different food categories. Also, Dhebibi *et al.* (2007) link product consumption with nutrient coefficients in a demand system. In principle, such demand models may be used for assessing the effects of SFA-taxation. This approach has been used in two Danish studies (Smed, Jensen & Denver, 2007; Jensen & Smed, 2007), which examine the effects of various tax change scenarios on the intake of different nutrients.

Santarossa & Mainland (2003) simulate the effect of introducing a tax to diminish consumers' fat intake. Based on their results, they argue that such an introduction will cut households' expenditure on the considered foods to such an extent that food producers will voluntarily reformulate food products towards more healthy alternatives.

Allais, Bertail & Nichele (2009) and Chouinard *et al.* (2007) simulate a fat tax on French and US households respectively with both studies concluding that it is regressive, i.e. the tax is especially burdensome for low income households who spend a higher share of their incomes on food, especially on unhealthy foods, hence foods more suitable as tax objects.

Gil, Angulo & Mtimet (2009) simulate the effects of taxing meat and subsidising FAV on dietary quality in Spain. Their analysis revealed a small impact on diet behaviour of this simulated intervention. However, the changes in the VAT rates were also quite small, especially for FAV (a drop in the VAT from 4% to 1%).

Cash, Sunding & Zilberman (2005) also simulate the effects of cheaper FAV, but on health outcome in the US. They estimated that a mean reduction of approximately 6.900 cases of CHD could be accomplished by a 1 percent cut in prices of FAV. Also, similar to their conclusion, one could argue that subsidising FAV would not suffer from the regressive nature of taxing foods high in SFAs and would offer the most benefits to the more disadvantaged consumers.

This argument is supported by paper 4 based on quantile regression, which revealed that consumers with the lowest consumption intensity of FAV were also the most price sensitive. Also, Beydoun *et al.* (2008) found that lower FAV prices were positively associated with improved dietary quality and were protective against obesity, particularly among the near poor.

Nordström & Thunström (2009) simulate the effects of tax reforms on the consumption of fibres and whole grains in Sweden. For the simulation, they estimate a demand system for grain products via two micro data sets: household expenditure data on soft bread from Statistics Sweden and household expenditure data on grain products from the marketing research firm GfK Sweden. The estimates are used to simulate the results of tax reforms designed to achieve the Swedish National Food Administration recommendations on fibre consumption. Their analysis revealed that a subsidy of 50 percent on wholesome bread and breakfast cereals would be necessary to achieve the national guidelines for recommended fibre intake. However, the analysis also revealed potential undesired by-products of the subsidy such as increased consumption of fat, sugar and salt. Consequently, they suggest simultaneously taxing unhealthy foods or nutrients to fund the subsidies and limit increases in unhealthy food consumption. A promising alternative or complement to a tax reform to encourage the consumption of fibres and whole grains is the use of nutrition labelling. The Danish wholegrain partnership (European Commission, 2008) introduced a new orange wholegrain logo in January 2009. Sales of whole grains with the recognisable orange logo is experiencing a massive growth among Danes, who are increasingly choosing crisp bread, wheat bread, cereals and flour with whole grains. Sales of products with the orange wholegrain logo rose by as much as 23% between weeks 1-36 in 2010 compared with the same period the year before, which is quite impressive when one considers the fact that the overall total growth in the market for grain products only increased by 2%, according to studies by the marketing research firm ACNielsen (Børsen, 2010; TV2, 2010).

It should be noted that research indicates that consumers often only read partial information pieces on food labels because of time constraints (Golan, Krissoff & Kuchler, 2007), indicating that identifying the most important piece of nutritional information may be difficult and too costly in terms of their opportunity cost of time. This could lead consumers to employ simple choice mechanisms neglecting the most relevant information in the process (Golan, Krissoff & Kuchler, 2007). Thus, these arguments and the empirical data from ACNielsen, support the hypothesis that the orange wholegrain logo and similar labelling can be successful as part of an eclectic choice strategy intended to help consumers handle the informational demands of multifaceted food choices, which indicates that consumers will benefit from the availability of simple and easy to interpret information. The absence of this could lead to cognitive strain<sup>8</sup>. Also, a study which tested the effectiveness of different nutrition labelling formats found that endorsement by health organisations strongly increased the labelling formats' credibility and that consumers needed significantly less time to evaluate simpler front-of-pack labelling (Feunekes *et al.*, 2008). This indicates that the endorsement of the wholegrain logo and its front-of-pack placement on products by health organisations could contribute to success. Note that there is no reason to discard the more thorough information labelling as long as it is complemented by easy to interpret information, and the more elaborate information does not increase the opportunity cost of time because it can be ignored, in the sense that the consumer is ensured a minimum standard.

---

<sup>8</sup> See Taylor (1975).

The above-mentioned studies are restricted in different ways.

First, evaluations are based on simulation using records of estimated price responsiveness of different foods at a much aggregated food category level for statistical reliability of results and computational tractability. A disadvantage of aggregation is that substitution within the aggregated food category levels is not modelled. One exception is the study by Griffith, Nesheim, O'Connell (2010) who use disaggregated household purchase level data, together with a discrete choice demand model, to estimate the effect of a tax on SFAs.

Secondly, a common feature of the studies is that the analysis is based on estimating food demand systems assuming a uniform statistical distribution of behavioural parameters across all individuals or households. A potential problem with this approach is that consumers might not be homogeneous in their response to price signals. Exceptions to this are the quantile based studies, e.g. Gustavsen & Rickertsen (2006, 2008) and paper 4. However, these studies do not allow an analysis based on a demand system of multiple demand equations.

Thirdly, empirical evidence exists that supports the fact that consumers show some asymmetry in their responses to price changes, e.g. price decreases tend to trigger larger demand responses than price increases for Danish milk drinkers (see paper 6). However, the above-mentioned studies do not take this effect into account.

Finally, evaluations based on simulations using records of the estimated price responsiveness of different foods still have disadvantages relative to evaluations based on real policy experiments even if the model underlying the simulation models heterogeneity and accounts for asymmetric price effects. This is true because the simulations are based on estimated price responsiveness modelling the short run consumption behaviour, and obviously not the effect of suddenly raising the baseline level which the prices revolve around.

The introduction of the Danish tax on SFAs in foods and its future evaluation should consequently prove very valuable in providing information about the effectiveness of fiscal measures in modifying consumer behaviour.

## Aims and objectives of thesis

The review has identified methodological problems and opportunities in the literature. The overall aim of the thesis is to:

1. Apply new methodologies that can identify problems or/and mitigate problems in previous applied policy evaluations.
2. Apply new methodologies that can better identify population heterogeneity in food consumption behaviour and consequently are better suited at targeting at risk groups and mitigate health disparities.
3. Derive empirical evidence of food consumption behaviour that can be used in policy decisions recognising the impact of population heterogeneity including development in inequality in health.

## Contribution to the literature

The following section gives an overview of contributions and results of each paper.

Paper 2 and 3 are methodological papers that deal with issues related to Paper 4.

Paper 2 entitled *A data problem in a popular dataset - Impact and solution* deals with a problem in the data, used in the thesis and delivered by the market research company GfK Denmark. The data problem is only an issue in paper 4. The implication of the data problem is that the values of prices in the data have to be estimated. An analysis of a naïve estimator formerly used is undertaken hereby assessing the seriousness of the problem. Finally a methodology based on the Expectation Maximisation (EM) algorithm is developed which gives much better results than the naïve estimator. The solution developed in this paper is used to cleanse the fruit and vegetables data used in paper 4.

Paper 3 deals with *estimation of multi-dimensional discrete stochastic variables*. The accuracy of the methodology used in paper 4 depends on how well the covariate distribution of the multi-dimensional discrete stochastic variables is estimated. Previous applications used a maximum likelihood estimator, however it is well known that when the number of observations is small relative to the number of parameters to estimate, it can be an advantage in finite sample applications to use a method that smoothes parameter estimates using a kernel function. Least square cross-validation is an example of such a smoothening approach and leads to differing asymptotic behaviour of the smoothening parameters depending on whether each variable is uniformly distributed or not. Also, smoothening parameters corresponding to uniformly distributed variables has positive probability of not converging to its optimal value when using least square cross-validation. This paper shows why this is the case and derives the asymptotic probability distribution of the smoothening parameter. Furthermore, a criterion function is suggested that solves this problem. The fact that the smoothening parameters now asymptotically converge to the optimal values effectively means that the associated estimator shrinks to the smallest variance unbiased frequency estimator and give support to the usability of smoothing methods in discrete variable applications.

## Education, information, health disparities and policy evaluation

It is well known that risk factors such as smoking, high body mass index, poor diet, etc. explain socio-economic differences in health to some extent (Arendt & Lauridsen, 2007). However, even after adjusting for such risk factors, socio-economic differences in health still exist, e.g. the study by Arendt & Lauridsen(2007) showed the existence of educational gradients in particular regarding health for women and occupational gradients for men. There could be several reasons why education could have an effect on health independent of common risk factors. One could be education-related environmental triggers. High education is less common among individuals possessing risk factors such as high body mass index (BMI), poor diet, smoking etc., which suggests education-related environmental triggers. Johnson *et al.* (2011) analysed the correlation between education and BMI using data on same-sex pairs of Danish twins including zygosity, height, weight, and education data and found that family influence was particularly important in linking high education and lower levels of obesity. Other reasons why education could have an independent effect on health might include its effect on individuals responsiveness to health related information and campaigns. One might

imagine that the educated segment of the population would be more likely to respond to a campaign message if decoding it requires formal schooling; if education helps develop strategies to change behaviour and if persons who are more responsive to authorities, self-select into further education. Finally, the educated segment of the population may be more likely to respond to a campaign message if the information is focused towards this particular segment of the population. This was the case for the Danish 6-a-day campaign (Ministry of Food, Agriculture and Fisheries, 2001). Great care should be taken when implementing information campaigns because of the governments' goal of eliminating health disparities together with lower educated individuals' higher marginal health return from complying with a health message, due to the social gradient, and because highly educated individuals might be more responsive to information campaigns. Also, because of the vast amount of resources put into information campaigns, it is of major importance to not only evaluate the effect of information campaigns on improving health, but also the extent to which they eliminate health disparities. Paper 4 examines the *determinants of development in inequality in fruit and vegetable consumption in Denmark*. Inequality in consumption of fruit and vegetables (FAV) in Denmark has increased in a period of extensive information campaigns, which had the purpose of increasing the consumption of FAV. The paper attempts to identify the reasons for this development. Quantile regression is used on a rich data set from the market research institute, GfK Denmark, for the analysis of the development in FAV consumption. The determinants of the development in FAV consumption are investigated by use of a matching methodology that enables comparisons of data samples with different covariate distributions. The study shows a negative relationship between the intensity of consumption and price sensitivity. Increased inequality in consumption over time can partly be attributed to this negative relationship coupled with increases in prices, but to a greater extent so that uneducated groups with a low consumption and low income groups are falling behind the more educated and financially well off segment of the population. The findings may be utilised to target specific sub-populations when designing policies aimed at increasing fruit and vegetable consumption. These findings highlight the advantages of quantitative methodologies that compare developments in samples and adjust for differences in covariate distributions.

As stated in the review most studies of hypothetical policies uses a uniform statistical distribution of behavioural parameters across all individuals or household and a potential problem with this approach is that consumers might not be homogeneous in their response to price signals.

A possible mitigation of this problem is to model the problem as done in the paper *consumption effects of a tax on saturated fat in foods* (paper 5). The purpose of this paper was to explain how one can forecast the effect of a proposed tax on saturated fat (SFAs) on the demand for butter and margarine. The tax is supposed to take effect in mid 2011 in Denmark. The tax was created to affect the consumption of SFAs and it is therefore of special interest to know if the population is price sensitive. Earlier studies simply focus on the mean price response, but this is unsatisfactory because most populations have a significant variation in their responses. Instead, the individual households' price response is estimated here by unlocking more information on the expected response to the tax reform. The main purpose was to examine if a tax on SFAs can improve the Danes' diet, while taking the possible population heterogeneity into account. The possibility of answering this question was obtained by taking advantage of a detailed panel dataset, and using a theoretical framework that excels by estimating household specific prices, and deals with missing prices by taking into account the preferences of households. Also, the estimation of the demand system at the household level puts less restrictive assumptions on the estimates than previous studies. A great heterogeneity within the population price responsiveness underlines the importance of the adopted approach and indicates that consumers should be treated

accordingly when considering the development of new food taxes. The analysis reveals that a likely outcome of the tax for the majority of the population is a decrease in the purchase of hard margarines and butter products, and that these will, to some extent, be substituted with healthier alternatives, such as spreadable margarine, minarine and vegetable oils. It is therefore expected that the tax reform will help to improve the Danes' fat intake profile, and consequently the health of the Danish population. Furthermore, the analysis reveals that the proposed SFAs tax is superior to a VAT on the considered food items in terms of modifying the consumers' food consumption behaviour in a desirable way. Paper 6 examines a hypothesis of an existence of *non-linearities in consumers' demand response to price changes* using Bayesian estimation techniques. The analysis brings the empirical demand analysis further by searching for potential non-linearities in the response to price changes. In general, the empirical analysis suggests that the demand for drinking milk responds significantly to price changes. But the analysis also shows some asymmetry in this response to price changes. In particular, price decreases tend to trigger larger demand response than price increases.

## Conclusion and further perspectives

Obesity rates are rising worldwide. Obesity affects productivity and the fiscal economy, which is increasingly important in an ageing European population, whilst it increases inequality in health.

Consequently, vast amounts of resources are invested into curtailing this development. Policy evaluation can be a cost effective measure in so far as it can help to guide resources in the most effective direction. This of course implicitly assumes that the undertaken policy evaluation is reliable.

However, this paper suggests that policy evaluation, in particular simple comparisons of pre-policy and post-policy surveys which are highly prevalent, is not always reliable, and provides evidence to support the assertion that it is important that policy evaluation, such as the evaluation of the 6-a-day campaign, does not merely measure the change in the sample of a quantity of interest, but also accounts for changes in the composition of distribution of socio-demographic variables such as education and income and accounts for the development in prices, if appropriate. Policy evaluations that do not account for such factors can give rise to misleading results.

The use of the evaluation results can also be limited if they do not reflect population heterogeneity and if reduction in health disparity is a policy goal in so far as policy implementation directed at specific target groups is hampered by the restricted information.

The methodology applied in paper 4 is interesting in that it introduces an attractive alternative to the unreliable evaluation consisting of a comparison of pre-policy and post-policy surveys and it also reflects population heterogeneity in the measure of interest, e.g. FAV. The methodology is therefore a useful tool in measuring developments in specific target groups such as individuals who do not eat much FAV, or eat too much SFAs.

Sometimes the survey analysed contains censored measurements, e.g. measurements of zero consumption, of the measure of interest, e.g. ice cream. This is often the case for consumption surveys conducted by national bureaus of statistics, especially for non-perishable foods such as ice cream. In this case, the methodology in paper 4 is not immediately applicable. However, the methodology is satisfactory in this case if algorithms used in censored quantile regression, such as those applied in Gustavsen & Rickertsen (2006, 2008), are incorporated into the methodology. The



introduction of such algorithms should prove straightforward, but will increase the computational burden of the methodology.

Another example of when the methodology applied in paper 4 could be beneficial is in studies which examine developments in specific target groups, such as individuals who eat too much SFAs, and when a simple measure of the total population effect of a tax change, i.e. the unconditional quantile, is of interest. Gustavsen & Rickertsen (2006, 2008) are interested in measuring this effect. However, as the authors point out, no exact method for the calculation of this exists. In the case of mean regression, finding the unconditional mean is straightforward when a regression model (conditional model) is available. One simply sums (or integrates) out the conditioning covariates of the expectation operator, i.e. utilise the law of the total probability (the law of iterated expectations). In the case of quantile regression, finding the unconditional value of the dependent variable is another matter; when a conditional model is available, receiving the unconditional quantile is non-trivial because no exact closed form formula similar to the law of iterated expectations exists. Gustavsen & Rickertsen (2006, 2008) use a closed form approximation to estimate the unconditional quantile after a tax increase. However, the accuracy of the closed form approximation depends on the shape of the cumulative distribution function (CDF) in different subsamples, as pointed out in Hansen (2009). An approximation that takes the shape of the CDF into account may produce more accurate results and this can be accomplished by using a methodology such as the one in paper 4.

Lauridsen *et al.* (2007) develop a model that integrates two methodologies so that income-related inequality in general health can be decomposed into contributions from socio-demographic characteristics to each of the dimensions defining general health. Lauridsen *et al.* find that health inequality is especially connected with low income and that the effects of socio-demographic determinants on different dimensions of health vary considerably. Based on their analysis, they argue that policy programmes, which aim to reduce income-related inequality in health, ought to be targeted at specific dimensions of health and at specific population groups, rather than being uniformly directed toward general health and the general population. A specific dimension of health could be the consumption of FAV, since low FAV intake is present in the top 8 leading risk factor causes of death in middle and high income countries (WHO, 2004). The reduction of inequality in health and FAV consumption is a policy goal of the 6-day-campaign. . Thus, the goal is in accordance with one of the recommendations of Lauridsen *et al.* (2007), in so far as the policy is aimed at a specific dimension of health. However, paper 4 finds that uneducated groups with a low consumption and low income groups are falling behind the more educated and financially well off segment of the population. The reason for this unfortunate development might be that the 6-a-day campaign is focused on the educated and “ready to change” part of the population (Ministry of Food, Agriculture and Fisheries, 2001). Thus, although the 6-a-day campaign is targeted at specific populations groups in accordance with another of the recommendations of Lauridsen *et al.* (2007), I believe that the campaign was targeted towards the wrong population groups, based on the evidence in paper 4 that the 6-a-day campaign and/or the health related information flow has increased inequality in FAV consumption.

A measure of success of a given policy is its ability to target the sub-population with the highest marginal health benefit. It is not obvious if a given policy has this trait. For example, Bere *et al.* (2005) study the effect of a fee-based School Fruit programme and conclude that the School Fruit Programme appears to increase the intake among the subscribers, but also increases inequality by increasing the gap in FAV consumption among subscribers and non-subscribers. One might fear that the response to a policy that decreases FAV prices would be that consumers with a high

intensity of FAV purchases would respond by increasing consumption, whilst households with a higher marginal health benefit from increased FAV consumption, i.e. low consuming groups, would not increase consumption to the same extent. This, apparently, has been the effect of the 6-a-day campaign. Paper 4, however, indicates that such fears might be unfounded, as low consuming groups have higher price responsiveness. Also, Bihan *et al.* (2010) uses a randomised controlled trial (RCT) methodology to investigate the effect of vouchers to increase FAV consumption in a deprived population. They identify financial difficulties in particular, as the perception of affordability of FAV is an obstacle, and that vouchers can increase the FAV consumption among individuals consuming particularly low amounts.

Policy evaluation could be improved if methodological and data restrictions are removed. The analysis in Paper 5 lifted one of these restrictions, i.e. a restriction on heterogeneity. The paper showed that a fat tax, such as the Danish tax on SFAs to be introduced mid-2011, has its merits, in that it was estimated to be better at modifying population consumption behaviour towards a more healthy diet, i.e. the consumption of products with higher PUFAs to SFAs ratio, than a VAT. However, studies such as this, which aim to forecast the consequences of new policies using historical data, have the property that the simulations are based on estimated price responsiveness to model short run consumption behaviour and obviously not the effect of suddenly raising the baseline level, which the prices revolve around. Also, empirical evidence suggests that consumers show some asymmetry in response to price changes, i.e. price decreases tend to trigger different demand responses than price increases (see paper 6). Hence, such simulation based studies should preferably use one estimate of price responsiveness when simulating tax increases and another estimate when simulating the introduction of subsidies, which is rarely if ever the case. The introduction of the Danish tax on SFAs in foods in mid-2011 and its future evaluation should consequently prove very valuable in providing information about the effectiveness of fiscal measures in modifying consumer behaviour.

Paper 5 clearly reveals the benefits of available data consisting of individual household level time series including food consumption information, i.e. dietary information and price, and socio-demographic status. The data, however, do have limitations consisting of possible inaccuracies in the measurement of food consumption. Although data from a market research institute, such as GfK Denmark, provide an objective measure of purchasing behaviour, and this type of data have less response bias relative to self-reported dietary measures (Hebert *et al.* (2008), Kristal *et al.* (1998)), the purchase diaries represent household-level purchase data and not the food consumption of individuals within households. However, studies have shown significant positive correlations between household-level food purchases and the dietary intake of individuals (Eyles *et al.*, 2010 and Ransley *et al.*, 2001).

Nayga (2008) argues that our understanding of dietary behaviour, and hence our ability to implement better health policies, largely depends on introducing innovative methods that can mitigate response bias in dietary measures. Such an innovative method, used by epidemiologists, is the emerging method of using biomarkers of dietary intake, as an objective measure of diet. It is, however, relatively expensive to collect biomarkers on people. Nayga therefore suggests that the *raison d'être* of biomarkers is to use them on a smaller subsample of a sample that represents a population of interest and for validation and calibration purposes only. Besides the necessity of a better understanding of dietary behaviour, it is also important to understand what constitutes an optimal diet. In the context of reducing CHD risk by reducing the intake of SFAs, Astrup *et al.* (2011) conclude that research is needed to find appropriate alternatives to specific foods to reduce CHD risk. Since the effect of particular foods on CHD risk cannot be predicted solely by their

content of total SFAs, it might be better to introduce a tax on foods containing SFAs where better alternatives exist, compared to introducing a uniform tax on SFAs in foods.

## References

Allais O., Bertail, P., and V. Nichele (2009). The Effects of a Fat Tax on French Households' Purchases: A Nutritional Approach. *American Journal of Agricultural Economics*, 92(1): 228-245.

Anderson P. M., Butcher K. F. & Levine P. B. (2003). Maternal employment and overweight children. *Journal of Health Economics*, Volume 22, Issue 3, May 2003, Pages 477-504.

Arendt J.N. and Lauridsen J. (2007). Health Inequalities. Do risk factors explain more of the social gradient in self-reported health when adjusting for baseline health? *European Journal of Public Health*, Vol. 18, No. 2, 131–137. doi:10.1093/eurpub/ckm096 Advance Access published on October 9, 2007.

Astrup, A.; Dyerberg J; Elwood, P.; Hermansen, K.; Hu, F. B.; Jakobsen, M. U.; Kok, F. J.; Krauss, R. M.; Lecerf, J. M.; LeGrand, P.; Nestel, P.; Risérus, U.; Sanders, T.; Sinclair, A.; Stender, S.; Tholstrup, T. and Willett, W. The role of reducing intakes of saturated fat in the prevention of cardiovascular disease: where does the evidence stand in 2010? *Am J Clin Nutr* 2011 ajcn.004622; First published online January 26, 2011. doi:10.3945/ajcn.110.004622.

Baser O. (2009) Too Much Ado about Instrumental Variable Approach: Is the Cure Worse than the Disease? Volume 12 , Number 8 , 2009 VALUE IN HEALTH.

Baylis K. & Dhar T. (2007). Effect of the Quebec Advertising Ban on Junk Food Expenditure. Working paper, November 2, 2007. <http://www.aae.wisc.edu/events/papers/AppEcon/2007/baylis.11.07.pdf>.

Bere E., Veierød M. B., Klepp K.(2005) The Norwegian School Fruit Programme: evaluating paid vs. no-cost subscriptions. *Preventive Medicine*. Volume 41, Issue 2, August 2005, Pages 463-470.

Berg C., Lappas G, Wolk A., Strandhagen E., Tore'n, K., Rosengren A., Thelle D, Lissner L.,(2009) Eating patterns and portion size associated with obesity in a Swedish population, *Appetite* 52: 21–26

Beydoun M. A., Powell L. M. and Wang Y. (2008) The association of fast food, fruit and vegetable prices with dietary intakes among US adults: Is there modification by family income? *Social Science & Medicine* Volume 66, Issue 11, June 2008, Pages 2218-2229.

Bhattacharya, Jay & Bundorf, M. Kate, 2009. The incidence of the healthcare costs of obesity, *Journal of Health Economics*, Elsevier, vol. 28(3), pages 649-658, May.

Bihan H., C. Mejean, S. Peneau, K. Castetbon, P. Galan, H. Le Clesiau, S. Hercberg (2010). Effect of vouchers to increase fruits and vegetables consumption in a deprived population: a randomized trial. EGEA 6<sup>th</sup> edition. Social and health benefits of balanced diet: The role of fruit and vegetables. European Commission, Brussels, Belgium – May 2010. [http://www.egeaconference.com/2010/Abstract\\_Book\\_EGEA\\_2010.pdf](http://www.egeaconference.com/2010/Abstract_Book_EGEA_2010.pdf)

Bosch G; Verbrugghe A; Hesta M; Holst JJ; van der Poel AFB; Janssens GPJ; Hendriks WH (2009). The effects of dietary fibre type on satiety-related hormones and voluntary food intake in dogs. *British Journal of Nutrition*, 2009; 102: 318-325.

Børsen(2010) Danskerne spiser mere fuldkorn. 21-10-2010.

[http://borsen.dk/nyheder/foedevareer/artikel/1/193843/danskerne\\_spiser\\_mere\\_fuldkorn.html](http://borsen.dk/nyheder/foedevareer/artikel/1/193843/danskerne_spiser_mere_fuldkorn.html)

Capacci, S., Mazzocchi, M., Five-a-day; a price to pay: an evaluation of the UK program impact accounting for market forces, *Journal of Health Economics* (2010), doi:10.1016/j.jhealeco.2010.10.006.

Cash, S. B., Sunding D. L., & Zilberman D. (2005). Fat Taxes and Thin Subsidies: Prices, Diet and Health Outcomes, *Food Economics: Acta Agriculturae Scandinavica*, sect. C, 2/3-4: 167-74.

Chapalain C. (2010). Food Industry Initiative in the framework of the Charter to promote healthy diet and physical activity in programmes and advertising, Catherine Chapalain, Director General, Association Nationale des Industries Alimentaires, France. Presented at: High Level Conference on Monitoring and Evaluation of EU and Member States' strategies on nutrition, overweight and obesity related health issues, 8 December 2010.

Charter D. (2010). Romanian Health Minister Attila Cseke to introduce 'fat tax' in battle against obesity. *The Times*, January 16, 2010. <http://www.timesonline.co.uk/tol/news/world/europe/article6990411.ece>

Chouinard, H.H., D.E. Davis, J.T. LaFrance, and J.M. Perlo. (2007). Fat taxes: big money for small change. *Forum for Health Economics & Policy*, 10(2): Berkeley Electronic Press.

Cobb-Clark, D. A. and Crossley, T. (2003), *Econometrics for Evaluations: An Introduction to Recent Developments*. *Economic Record*, 79: 491–511. doi: 10.1111/j.1475-4932.2003.00148.x

Cutler, D. M., Glaeser, E. L., & Shapiro, J. M. (2003). Why have Americans become more obese? *Journal of Economic Perspectives*, 17, 93-118.

DaeHwan Kim, John Paul Leigh. Estimating the Effects of Wages on Obesity. *Journal of Occupational and Environmental Medicine*, 2010; 52 (5): 495 DOI: [10.1097/JOM.0b013e3181dbc867](https://doi.org/10.1097/JOM.0b013e3181dbc867)

Dam, Thea, Jensen, Jørgen Dejgaard and Kærgård, Niels (2008)'Obesity, social inequality and economic rationality: An overview', Food Economics - Acta Agriculturae Scandinavica, Section C,5:3,124 — 137.

Dhebibi B., Gil J.M. & Angulo A.M. (2007) Nutrient Effects in Consumer Demand Systems: Evidence from Panel Data, Food Economics, vol. 4, pp. 89-102.

Dor A.,Ferguson C., Langwith C., Tan E. (2010),

Research Report: A Heavy Burden: The Individual Costs of Being Overweight and Obese in the United States. The George Washington University School of Public Health and Health Services. Department of Health Policy. September 21, 2010.

European Food Safety Authority (2009). Scientific Opinion of the Panel of Dietetic Products, Nutrition and Allergies on a request from the European Commission related to dietary reference values for fat. The EFSA Journal (2009), 1-97.

Available at:

[http://www.efsa.europa.eu/en/calls/documents/nda\\_op\\_drv\\_fat\\_draft\\_en\\_released%20for%20consultation.pdf](http://www.efsa.europa.eu/en/calls/documents/nda_op_drv_fat_draft_en_released%20for%20consultation.pdf).

Accessed:11.3.2009:

European Commission (2005) Green Paper of 8 December 2005. *"Promoting healthy diets and physical activity: a European dimension for the prevention of overweight, obesity and chronic diseases"* (COM (2005) 637 final).

European Commission (2007). WHITE PAPER ON A Strategy for Europe on Nutrition, Overweight and Obesity related health issues. Brussels, 30.5.2007. Available at

[http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/documents/nutrition\\_wp\\_en.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/nutrition_wp_en.pdf)

European Commission (2008) Working Paper on Public-Private Partnerships for Health,

NOVEMBER 2008. Accessed on 16.10.10:

[http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/documents/ev20081028\\_wp\\_en.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/ev20081028_wp_en.pdf)

EU platform on diet, physical activity and health (2010). Diet, Physical Activity and Health -a European Platform for Action. Presentation by Ida Enghave, Danish Heart Foundation, Andrzej Wojtyła, Polish Chief Sanitary Inspector, *et al.*, 11. February 2010. [http://ec.europa.eu/health/nutrition\\_physical\\_activity/docs/ev20100211\\_mi\\_en.pdf](http://ec.europa.eu/health/nutrition_physical_activity/docs/ev20100211_mi_en.pdf)

Eyles HC, Jiang Y, Mhurchu C. N. (2010). Use of household supermarket sales data to estimate nutrient intakes: a comparison with repeat 24-hour dietary recalls. J Am Diet Assoc 2010;110:106–10.

Feunekes G. I.J., Gortemaker I. A., Willems A. A., Lion R., van den Kommer M. (2008). Front-of-pack nutrition labelling: Testing effectiveness of different nutrition labelling formats front-of-pack in four European countries. Appetite 50 (2008) 57–70.

Finkelstein, E. A., Ruhm, C. J., & Kosa, K. M. (2005). Economic causes and consequences of obesity. Annual Review of Public Health, 26, 239-257.

French S. A., Harnack L. & Jeffery R. W. (2001). Fast food restaurant use – dietary behavioural and demographic correlates in a sample of adult women. *International Journal of Obesity*, 24: 1353-1359.

Golan E., Krissoff B. & Kuchler F. (2007). Do Food Labels Make a Difference?. . . *Sometimes*

AMBER WAVES VOLUME 5, ISSUE 5, November 2007.

<http://www.ers.usda.gov/AmberWaves/November07/PDF/FoodLabels.pdf>

Gil, J. M., Angulo, A.M., & Mtimet N. (2009). The impact of Price Policies to improve Dietary Quality in Spain. Paper presented at the 2009 International Association of Agricultural Economists Conference: Diet and Obesity: Role of Prices and Policies.

Griffith R., Nesheim L., O'Connell M. (2010). Comparing pass-through of excise and ad valorem taxes in differentiated product oligopoly with multi-product firms: empirical estimates using micro data. Working Paper, Institute for Fiscal Studies: London, UK.

Gordon R, McDermott L, Stead M, and Angus K. (2006) The effectiveness of social marketing interventions for health improvement: What's the evidence? *Public Health* 2006;120; 1133-1139.

Government /Regeringen (2002) Sund hele livet - *de nationale mål og strategier for folkesundheden 2002-10*.

Regeringen September 2002. Udgivet: September 2002. Udgivet af Indenrigs- og Sundhedsministeriet, Slotsholmsgade 10-12, 1216 København K. ISBN 87-90227-67-0, trykt version, ISBN 87-90227-69-7, elektronisk version.

Grootendorst P. (2007). A review of instrumental variables estimation of treatment effects in the applied health sciences. Health Services and Outcomes Research Methodology, Volume 7, Numbers 3-4, 159-179, DOI: 10.1007/s10742-007-0023-6

Gustavsen, G. W., and Rickertsen K. (2006). "A Censored Quantile Regression Analysis of Vegetable Demand: The Effects of Changes in Prices and Total Expenditure." *Canadian Journal of Agricultural Economics* 54:631-45.

Gustavsen G.W., D. Jolliffe & Rickertsen K. (2008). "Censored Quantile Regression and Purchases of Ice Cream," *Food Economics*, 5: 152-163.

Hansen A. H. (2009). Forecasting the consumption effect of taxing foods containing saturated fat. Paper prepared for presentation at the 113th EAAE Seminar "A resilient European food industry and food chain in a challenging world", Chania, Crete, Greece. September 3 - 6, 2009.

Hasselbalch A L; Heitmann B L; Kyvik K O; Sørensen T I A (2010). Associations between dietary intake and body fat independent of genetic and familial environmental background. *Int J Obes* 2010 34: 892-898.

Havas S. et al. (1995) 5 A Day for Better Health-Nine Community Research Projects

To Increase Fruit and Vegetable Consumption. *Public Health Reports*. January-February 1995, Vol. 110, No. 1 71.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1382077/pdf/pubhealthrep00056-0070.pdf>.

- Heckman, J. (2008). Econometric causality. *International statistical review* 76: 1-27.
- Hebert J. R., Hurley T. G., Peterson K. E., Resnicow K., Thompson F. E., Yaroch A. L., Ehlers M., Midthune D., Williams G. C., Greene G. W., and Nebeling L. (2008) Social Desirability Trait Influences on Self-Reported Dietary Measures among Diverse Participants in a Multicenter Multiple Risk Factor Trial. *J. Nutr.* 138: 226S–234S, 2008.
- Herman P. & Polivy J. (2003). Dieting as an Exercise in Behavioural Economics. In Loewenstein G., Read D. & Baumeister R. F. (ed.), *Time and Decision. Economic and Psychological perspectives on intertemporal choice*. ISBN 0-87154-549-7. 459-491.
- Ho D.E. , Imai K., King G., Stuart E. A.(2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* (2007) 15 (3): 199-236. doi: 10.1093/pan/mpi013
- Huang K.S. & Lin B.H. (2000) Estimation of Food Demand and Nutrient Elasticities from Household Survey Data, Technical Bulletin no. 1887, Economic Research Service, US Department of Agriculture, Washington DC, 30 pp.
- IASO/IOTF (2010). International Obesity Taskforce. <http://www.iaso.org/iotf/obesity/obesitytheglobalepidemic/>
- Jakab Z. (2010). The WHO's Progress report on the implementation of the European Charter on counteracting obesity. Zsuzsanna Jakab, Regional Director, WHO Regional Office for Europe. Presented at: High Level Conference on Monitoring and Evaluation of EU and Member States' strategies on nutrition, overweight and obesity related health issues, 8 December 2010.
- Jensen J.D. & Smed S. (2007) Cost-effective design of economic instruments in nutrition policy, *International Journal of Behavioral Nutrition and Physical Activity* 2007, **4**:10 doi:10.1186/1479-5868-4-10
- Jones A. (2008). Identification of treatment effects in health economics. *Health Economics* 16: 1127-31.
- Johnson W, Kyvik KO, Skytthe A, Deary IJ, Sørensen TIA. ( 2011). Education modifies genetic and environmental influences on BMI. *PLoS ONE*, 2011;6:e16290.
- Koc A., Dolekoblü T. & Erdal Ertuk Y. (2001) Household Vegetable Oil and Butter Demand in Turkey: A Three Stage Budgeting and AIDS with Adjusted Unit Value, paper presented at the 71<sup>st</sup> EAAE Seminar: The Food Consumer in the Early 21<sup>st</sup> Century, Zaragoza (Spain), April 19-20, 2001
- Kristal A. R., Andrilla C. H. A., Koepsell T. D., Diehr P. D (1998) Dietary assessment instruments are susceptible to intervention-associated response set bias. *J Am Diet Assoc* 1998;98:40–3.
- LAAKSONEN D. E., LAKKA H., SALONEN J. T., NISKANEN L. K., RAURAMAA R., LAKKA T. A. (2002). Low Levels of Leisure-Time Physical Activity and Cardiorespiratory Fitness Predict Development of the Metabolic

Syndrome. *Diabetes Care* 25:1612–1618, 2002

L'Abbe M. R., Stender S., Skeaff M., Ghafoorunissa & Tavella M. (2009). Approaches to removing trans fats from the food supply in industrialized and developing countries." *Eur J Clin Nutr* 63, no. S2(2009): S50-S67.

LAIBSON D. (1997). GOLDEN EGGS AND HYPERBOLIC DISCOUNTING. *The Quarterly Journal of Economics*, May 1997.

Lakdawalla, D., & Philipson, T. (2002). The growth in obesity and technological change: A theoretical and empirical examination. NBER Working Paper 8946.

Lakdawalla, D. and Philipson, T. (2006). Economics of obesity. In Jones, A. M. (ed.), *The Elgar Companion to Health Economics*. Northampton, MA: Edward Elgar, 72-82.

Lang T, Rayner G & Kaelin E.(2006) The food industry, diet, physical activity and health: a review of reported commitments and practice of 25 of the world's largest food companies. Working paper Centre for Food Policy, City University, April 2006.

<http://www.city.ac.uk/news/press/The%20Food%20Industry%20Diet%20Physical%20Activity%20and%20Health.pdf>

Lauridsen, J., Christiansen, T., Gundgaard, J., Häkkinen, U. and Sintonen, H. (2007), Decomposition of health inequality by determinants and dimensions. *Health Economics*, 16: 97–102. doi: 10.1002/hec.1145

LIVSFS 2005:9. [http://www.slv.se/upload/dokument/Nyheter/2005/2005\\_9.pdf](http://www.slv.se/upload/dokument/Nyheter/2005/2005_9.pdf)

Malik, V.S., Schulze, M.B. & Hu, F.B. Intake of sugar-sweetened beverages and weight gain: a systematic review. *Am J Clin Nutr* **84**, 274-288 (2006).

Mancino, L., Kuchler, F. & Leibtag, E. Getting consumers to eat more whole-grains: the role of policy, information and good manufacturers. *Food Policy* **33**, 489-496 (2008).

Martin S. S. (2005). From Poverty to Obesity: Exploration of Food Choice Constraint Model and the Impact of energy-dense Food Tax, *American Economist*, 49(2): 78-86

McClure S. M., Laibson D. I., Loewenstein G., and Cohen J. D.(2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306, 503–7.

Ministry of Food, Agriculture and Fisheries/Fødevaredirektoratet (2001) *6 om dagen – Overordnet kampagnestrategi*. Ref.: LSL/ETR. J.nr.: KOMS 3520-1/00. 10. Januar 2001. Fødevaredirektoratet

Modigliani F. (1975). The life-cycle hypothesis of saving twenty years later. In Parkin M.(ed.), *Contemporary Issues in*



*Economics*. Manchester. Manchester University Press, 2–35.

Mozaffarian D., Micha R., Wallace S.(2010) Effects on Coronary Heart Disease of Increasing Polyunsaturated Fat in Place of Saturated Fat: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PLOS Medicine*. Volume 7. Issue 3, pp 1-10.

Nayga R. M. (2008). Nutrition, obesity and health: policies and economic research challenges. *Eur Rev Agric Econ* (2008) 35 (3): 281-302. doi: 10.1093/erae/jbn013

Nichele V. (2003) Health Information and Food Demand in France, chapter 8 in Chern W.S. & Rickertsen K. (eds.) *Health, Nutrition and Food Demand*, CAB International

Niederdeppe J. & Frosch Dominick L. (2009). News Coverage and Sales of Products with Trans Fat Effects Before and After Changes in Federal Labeling Policy. *Am J Prev Med* 2009;36(5):395–401.

Nestle, Marion (2002), *Food Politics*, University of California Press.

Nilsson AC; Östman EM; Holst JJ; Björck IM(2008). Including indigestible carbohydrates in the evening meal of healthy subjects improves glucose tolerance, lowers inflammatory markers, and increases satiety after a subsequent standardized breakfast. *Journal of Nutrition* 2008; 138: 732-39.

Nilsson AC; Östman EM; Knudsen KE; Holst JJ; Björck IM(2010) A cereal-based evening meal rich in indigestible carbohydrates increases plasma butyrate the next morning. *J Nutr*. 2010. Sept 1, 2010.

Nordström, J. and L. Thunström. (2009). The Impact of tax reforms designed to encourage healthier grain consumption. *Journal of Health Economics*,28: 622-634.

Platform Charter (2005). EU Platform on Diet, Physical Activity and Health. DIET, PHYSICAL ACTIVITY AND HEALTH - A EUROPEAN PLATFORM FOR ACTION. 15 March 2005. Available at [http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/platform/docs/platform\\_charter.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/platform/docs/platform_charter.pdf). Accessed on 15 Marts 2010.

Polmark technical report of workpackage 5 (2010). Stakeholder views on policy options for marketing food and beverages to children. Cross-national report. January 2010.  
<http://www.polmarkproject.net/documents/StakeholderViews-Cross-nationalReport.pdf>

Puska P. (2002). Successful prevention of non-communicable diseases: 25 year experiences with North Karelia Project in Finland *Public Health Medicine* 2002; 4(1):5-7.  
[http://www.who.int/chp/media/en/north\\_karelia\\_successful\\_ncd\\_prevention.pdf](http://www.who.int/chp/media/en/north_karelia_successful_ncd_prevention.pdf)

Project Europe 2030 on the Challenge of Demography (2010). *Population and Development Review*, 36: 647–650.  
doi: 10.1111/j.1728-4457.2010.00350.x

[http://www.reflectiongroup.eu/wp-content/uploads/2010/05/reflection\\_en\\_web.pdf](http://www.reflectiongroup.eu/wp-content/uploads/2010/05/reflection_en_web.pdf)

Ransley J. K., Donnelly J. K., Khara T. N., Botham H., Arnot H., Greenwood D. C. and Cade J. E. (2001). The use of supermarket till receipts to determine the fat and energy intakes in a UK population. *Public Health Nutr* 2001;4:1279–86.

Rickertsen K. & von Cramon-Taubadel S. (2003). Health, Nutrition and Demand for Food: a European Perspective. In Chern W. S. & Rickertsen K. (ed.), *Health, Nutrition and Food Demand*. CABI Publishing, 33-52.

Roskam A. & Kunst A. E. Overview of inequalities in overweight and obesity across Europe. Chapter 23 in *Tackling Health Inequalities in Europe: an Integrated Approach*. Eurothine Final Report. Rotterdam: Erasmus University Medical Centre, 2007.

[http://mgzlx4.erasmusmc.nl/eurothine/uploads/eurothine\\_final\\_report\\_complete.zip](http://mgzlx4.erasmusmc.nl/eurothine/uploads/eurothine_final_report_complete.zip), accessed 10 Jan 2010.

Santarossa J. M. & Mainland D. D. (2003) Employing an environmental taxation mechanism to reduce fat intake. In Chern W. S. & Rickertsen K. (ed.), *Health, Nutrition and Food Demand*. CABI Publishing, 223-246.

Schreyögg J. and Grabka, M. M. (2010). Copayments for ambulatory care in Germany: a natural experiment using a difference-in-difference approach. *The European Journal of Health Economics*. Volume 11, Number 3, 331-341, DOI: 10.1007/s10198-009-0179-9

Smed S., Jensen J.D. & Denver S. (2007) Socio-economic characteristics and the effect of taxation as a health policy instrument, *Food Policy*, vol. 32, pp. 624-639

Stables *et al.* (2002). Changes in vegetable and fruit consumption and awareness among US adults: Results of the 1991 and 1997 5 A Day for Better Health Program surveys. *J Am Diet Assoc*. 2002;102:809–817.

Stender S., Astrup A., Dyerberg J. (2009). What Went In When Trans Went Out?  
*N Engl J Med* 2009; 361:314-316 July 16, 2009

Stewart H., Blisard N and Jolliffe D (2003). Do income constraints inhibit spending on fruit and vegetables among low-income households? *Journal of agricultural and resource economics* 28(3):465-480.

Sturm R., Powell L. M., Chiqui J. F. & Chaloupka F. J. (2010). Soda Taxes, Soft Drink Consumption, And Children's Body Mass Index. *Health Aff* April 2010.

Sørensen J., Sørensen J. B., Skovgaard Thomas., Bredahl T., Puggaard L. (2010). Exercise on prescription: changes in physical activity and health-related quality of life in five Danish programmes. *European Journal of Public Health*, Vol. 21, No. 1, 56–62.

doi:10.1093/eurpub/ckq003 Advance Access published on 5 April 2010

Sørensen Thorkild I. A. (2009). Conference on "Multidisciplinary approaches to nutritional problems". Symposium on "Diabetes and health". Challenges in the study of causation of obesity. Proceedings of the Nutrition Society 2009: 68 (1): 43-54.

Taylor J. (2008). Evaluation of the 'Girls on the Move' Programme: Summary Report  
A Report to the Robertson Trust and the Scottish Government. Department of Sports Studies  
University of Stirling August 2008. <http://www.youthscotland.org.uk/projects/girls-on-the-move.htm>

Taylor R. N.(1975). PSYCHOLOGICAL DETERMINANTS OF BOUNDED RATIONALITY: IMPLICATIONS FOR DECISION-MAKING STRATEGIES. Decision Sciences, Volume 6, Issue 3, pages 409–429, July 1975. DOI: 10.1111/j.1540-5915.1975.tb01031.x

The Danish Ministry of Taxation (2009,I). Lovforslag om afgift på mættet fedt, 18.08.2009. Available at <http://www.skm.dk/public/dokumenter/lovstof/Lovforslag%20maettet%20fedt.pdf>. Accessed on 15.3.2010.

The Danish Ministry of Taxation (2009,b). Danish Tax Reform 2010 - Paper to the OECD WP 2 meeting November 2009 by The Danish Ministry of Taxation. Available at [http://www.skm.dk/public/dokumenter/engelsk/Danish%20Tax%20Reform\\_2010.pdf](http://www.skm.dk/public/dokumenter/engelsk/Danish%20Tax%20Reform_2010.pdf). Accessed 15.3.2010.

The Danish Ministry of Taxation (2011, a). L 111 Forslag til lov om afgift af mættet fedt i visse fødevarer (fedtafgiftsloven)  
<https://www.retsinformation.dk/Forms/R0710.aspx?id=135445>

The Danish Ministry of Taxation (2011, b). Tidsplan for L 111 Forslag til lov om afgift af mættet fedt i visse fødevarer (fedtafgiftsloven). <http://www.ft.dk/samling/20101/lovforslag/1111/tidsplan.htm>  
TV2(2010) Danskerne spiser mere fuldkorn. Onsdag d. 7. jul. 2010  
<http://nyhederne-dyn.tv2.dk/article.php/id-31994300:danskerne-spiser-mere-fuldkorn.html>

US Department of Health and Human Services (2000). *Healthy People 2010* (Conference edition, in two volumes). Washington, DC: January 2000.

U.S. Department of Health and Human Services(2003). Statistics Related to Overweight and obesity. NIH Publication, No. 03-4158, Washington DC.

U.S Department of Health and Human Services (2004). Conquering Obesity: The U.S Approach to Combating this National Health Crisis. September 15, 2004.

Accessed on 17.11.10: <http://www.hhs.gov/asl/testify/t040915.html>.

van Baal PHM, Polder JJ, de Wit GA, Hoogenveen RT, Feenstra TL, et al. (2008) Lifetime Medical Costs of Obesity: Prevention No Cure for Increasing Health Expenditure. *PLoS Med* 5(2): e29. doi:10.1371/journal.pmed.0050029

Vermeer W. M., Steenhuis, & Seidell J. C. (2009a) From the point-of-purchase perspective: A qualitative study of the feasibility of interventions aimed at portion-size, *Health Policy* 90: 73–80

Vermeer W.M, Alting E., Steenhuis I. & Seidell J. C. (2009b) Value for money or making the healthy choice: the impact of proportional pricing on consumers' portion size choices, *Eur J Public Health*. 20(1):65-9.

Wakefield M.A., Loken B., Hornik R. C. (2010). Use of mass media campaigns to change health behavior. *The Lancet*, Volume 376, Issue 9748, 9 October 2010-15 October 2010, Pages 1261-1271

Walker, K.Z., Woods, J.L., Rickard, C.A. & Wong, C.K. Product variety in Australian snacks and drinks: how can the consumer make a healthy choice? *Public Health Nutr* **11**, 1046-1053 (2008).

Wansink B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of nutrition*, 24:455-479.

WHO(1990). Diet, nutrition and the prevention of chronic diseases. Report of a WHO Study Group. 108 (World Health Organisation, Geneva, 1990).

WHO (2000). Obesity: Preventing and managing the global epidemic report of a WHO consultation. WHO Technical Report Series, No. 894.

<http://www.iaso.org/policy/aboutobesity/>

WHO (2003, a). Diet, nutrition and the prevention of chronic diseases: report of the joint WHO/FAO expert consultation.

WHO/FAO (2003) Expert Report on Diet, Nutrition and the Prevention of Chronic Diseases, Technical Report Series 916, 2003, WHO/FAO.

WHO (2003, b). Obesity and Overweight. Global Strategy on Diet, Physical Activity and Health. Available at <http://www.who.int/dietphysicalactivity/media/en/gsf Obesity.pdf>. Accessed on 15 Marts 2010.

WHO (2004a) GLOBAL HEALTH RISKS. Mortality and burden of disease attributable to selected major risks.

[http://www.who.int/healthinfo/global\\_burden\\_disease/GlobalHealthRisks\\_report\\_full.pdf](http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf)

WHO (2006). Fact sheet N°311, September 2006. Available at:  
<http://www.who.int/mediacentre/factsheets/fs311/en/index.html>. Accessed on 15 Marts 2010.

WHO(2008). WHO European Action Plan for Food and Nutrition Policy. 2007-2012.

Young L. & Swinburn B. (2002). Impact of the *Pick the Tick* food information programme on the salt content of food in New Zealand. *Health Promot. Int.* (2002) 17 (1): 13-19. doi: 10.1093/heapro/17.1.13

Yen S.T. & Chern W.S. (1992) Flexible Demand Systems with Serially Correlated Errors: Fats and Oil Consumption in the United States, *American Journal of Agricultural Economics*, vol. 74(3), pp. 689-696

Østbye T, Dement JM, Krause KM. Obesity and workers' compensation: results from the Duke Health and Safety Surveillance System. Arch Intern Med. 2007 Apr 23;167(8):766-73. doi:10.1001/archinte.167.8.766

# Paper 2:

## A data problem in a popular dataset

### Impact and solution

---

Aslak Hedemann Hansen  
University of Copenhagen, Institute of food and resource economics

#### **Abstract**

This paper deals with a problem in the data delivered by the market research company GfK Denmark. The data problem is only an issue in paper 4 which analyses fruit and vegetable demand. The implication of the data problem is that the values of prices in the data have to be estimated. For some groceries the problem is absent, making it possible to create datasets with and without the problem. An analysis of a naïve estimator formerly used is undertaken to assess the seriousness of the problem. Finally, a methodology based on the Expectation Maximisation (EM) algorithm is developed which gives much better results than the naïve estimator. The solution developed in this paper is used to cleanse the fruit and vegetables data used in paper 4.

#### **1. Introduction**

The data, GfK Consumer Tracking data, used in the thesis is delivered by the market research company GfK Denmark. This paper is concerned with a problem in this data. The market research company has managed the data in such a way that the exact amount of food bought and the price paid is unknown in some cases.

It is important for the market research company, GfK Denmark, that they have representative data for their analyses. Since some types of households are underrepresented, a correction is needed. The imbalance was earlier corrected by doing multiple copies on underrepresented households. This type of data management is not a problem since these extra observations can be deleted. From 2001, the correction was instead done by weighting each observation; underrepresented households were then given a weight greater than overrepresented households. The data problem arises because, after weighting each observation, e.g. multiplying the quantity bought by the weight, the number was rounded to the nearest integer, thereby making it impossible to recover the exact quantity bought with absolute certainty in some cases.

The aim of this paper is to assess a formerly used method for handling this problem and to develop a better method that can provide a better estimate of the true quantity, price and value of the food purchased. The new methodology is based on the Expectation Maximisation (EM) algorithm and gives much better results than the previously used naïve estimator.

#### **A short overview of the EM algorithm**

The name, Expectation Maximization (EM) algorithm was coined in a paper by Dempster, Laird & Rubin (1977). Its

named is derived from the fact that the algorithm iterates between different steps where one step can often be considered as using the expectation operator, whilst another step maximises a function. Their paper reviewed many (at the time) recent statistical applications that all used the same type of algorithm for computing maximum likelihood estimates. The authors also showed some theoretical results relating to its convergence. Wu (1983) is also often cited regarding results on the convergence of the EM algorithm. Actually, the EM algorithm, or rather the very closely related concept of self-consistency, can be tracked all the way back to Fischer (Efron, 1982). An interesting early application in this new early wave of applications of the EM algorithms is the Efron (1967) application to incomplete observations, closely related to the methodology used in this paper, in which he shows how the self-consistency property can be used to derive the Kaplan & Meier (1958) product limit estimator for right censored duration data. Today, owing to its generality, the EM algorithm is used in a very broad range of applications including incomplete observations and advanced statistical models such as Monte Carlo Markov Chain based models.

Section 2 starts with an analysis of a formerly used estimator, here named the naïve estimator to exemplify the magnitude of the problem. A comparison is made of the true values and the values based on the naïve approach used in earlier published work. The conclusion is that there is a significant bias in the estimates if the naïve estimator is used, which stresses the importance of developing a method that can give values closer to the true ones. Section 3 takes a closer look at the structure of the problem and starts the development of a method that can provide better results than the naïve method. Section 4 suggests an estimator based on the Expectation Maximisation (EM) algorithm for solving the problem. Section 5 demonstrates that a previous suggestion for smoothing the estimates based on the EM algorithm is in fact no improvement at all.

## 2. The naïve approach

Let  $q$  and  $v$  be the true, but unknown quantity and value of a good bought by a household,  $w$  the weight and  $z_q = \text{Int}(q * w)$  the known, adjusted and rounded<sup>9</sup> quantity,  $z_v = \text{Int}(v * w)$  the known, adjusted and rounded value. Some publications then implicitly used the following estimators for quantity, value and price:

$$(2.1) \quad \hat{\theta}_q = \frac{z_q}{w} = \frac{\text{Int}(q * w)}{w}, \quad \hat{\theta}_v = \frac{z_v}{w} = \frac{\text{Int}(v * w)}{w}, \quad \hat{\theta}_p = \frac{z_v / w}{z_q / w} = \frac{\text{Int}(v * w) / w}{\text{Int}(q * w) / w} = \frac{\text{Int}(v * w)}{\text{Int}(q * w)}$$

If the true quantity, value and price were known, it would be possible to find the bias by using this estimator:

$$(2.2) \quad \text{bias}_q = q - \hat{\theta}_q, \quad \text{bias}_v = v - \hat{\theta}_v, \quad \text{bias}_p = p - \hat{\theta}_p$$

Since the bias is much more harmful in cases where the quantity is low, the percentage bias is calculated:

$$(2.3) \quad \text{bias}\%_q = \left( \frac{q - \hat{\theta}_q}{q} \right) * 100\%, \quad \text{bias}\%_v = \left( \frac{v - \hat{\theta}_v}{v} \right) * 100\%, \quad \text{bias}\%_p = \left( \frac{p - \hat{\theta}_p}{p} \right) * 100\%$$

---

<sup>9</sup> In section 3, an exact definition of the function  $\text{Int}()$  is given.

In the data, it is in fact possible to find the exact quantity of milk and therefore milk is used to calculate this bias and can serve as an example of how biased estimates of quantities of foods bought in similar quantities will be.

### The case of milk

Milk is a good case for investigating the bias since the true quantity can be found by using the naïve approach. This is the case because milk is measured in grams and the weights only have 3 decimals. As an example, the case where a household buys one litre of milk is used. In the data this is measured as 1000 grams, also in this example it is assumed that the weight is  $w=0.611$ . Then, the data  $z_q = 611$  and  $w = 0.611$  are given, which means that the naïve approach gives:

$$\frac{1000}{0.611} = 1636.661$$

But, what if the quantity was measured in litres instead of grams? Then the calculation would be:

$$\frac{1}{0.611} = 1.636661 \text{ and } 1.636661 \times 1000 = 1636.661$$

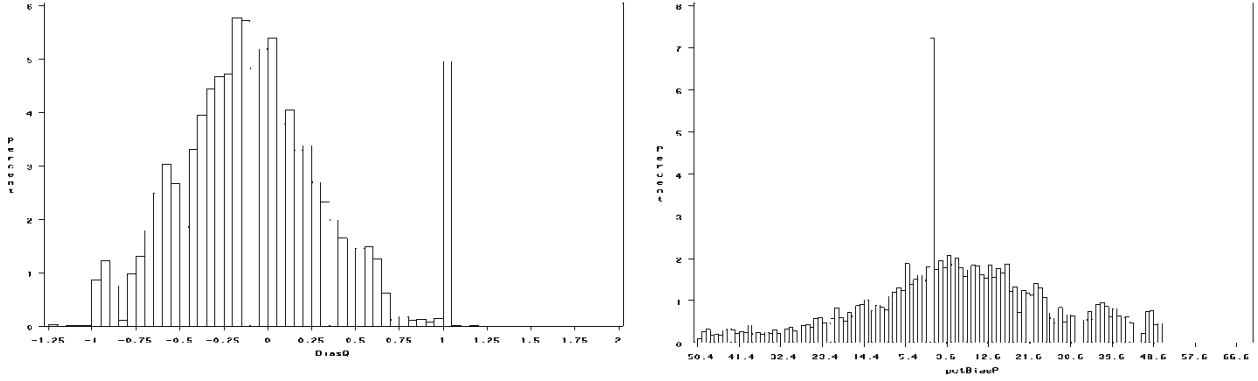
Since households usually only buy a few litres of milk and the greatest percentage bias exists when the quantity bought is low, the case of milk can be seen as a worst case scenario, but similar to other types of food bought in small numbers such as watermelons and coconuts.

### Assessment of the naïve estimator

Figure 1 shows the magnitude of the problem when using the naïve estimator. A total of 611.861 milk purchases are used. The left panel of figure 1 is the distribution of the bias of the naïve estimator calculated according to (2.2) and the right panel is the distribution of the %-bias of price calculated by (2.3). In the left panel of figure 1, approximately 5 percent of the shopping trips have a bias of 1, which seems a little peculiar when considering the rest of the distribution. Most of these shopping trips consisted of a purchase of one litre of milk and would equal 0 when using the naïve estimator in (2.1) if the weight has values:  $w < 0.5$ , hence a bias of 1 for all the observations. All of the observations with values equaling 0 are in fact known to have true values equaling 1 because there is no weight as small as 0.25. All of these observations have been corrected in the right panel of figure 1 and that is why there is such a large percentage with %-bias of zero. Figure 1 shows that the naïve estimator is in fact seriously biased and this will in general hold for any food group - not just for milk. The bias is especially problematic for food groups bought in small numbers such as milk, watermelons, coconuts etc. In the right panel, it can be seen that prices are also significantly biased when using the naïve estimator. Models involving prices as important components can then be expected to be significantly biased.



**Figure 1** Distribution of Bias for the naïve estimator for quantity (left) and (right)



The left panel is produced using formula (2.2), i.e. the bias of the quantity and the right is obtained from (2.3), i.e. the percentage bias in price.

### 3. The structure of the problem

The reason why we cannot recover the true values is because the rounding function used in the data treatment is not injective.

We have unobservable values in the data material because of the relationship:

$$f : q \rightarrow z$$

Where the function  $f$  and  $z$  is known but  $q$  is unknown. The existence of an inverse function  $f^{-1}$  is not possible because  $f$  is not injective. Rather, for each value of  $z$ ,  $q$  is only known to be belonging to a set with possible more than one element,  $q \in g(z)$ . Here  $g$  is a correspondence.

The exact value of  $q$  is unknown, but is known to be an integer. Instead,  $w$  and  $z$  are known and the function  $f$  is known.  $f$  is defined by:

$$z = f(qw) = f(a + b) = \begin{cases} a & \text{if } b < 0.5 \\ a+1 & \text{if } b \geq 0.5 \end{cases}$$

Where  $q$  and  $a$  are integers,  $w$  and  $b$  are real numbers,  $0 \leq b < 1$  and  $qw = a + b$ . Therefore  $q$  is an integer belonging to the interval given by the correspondence

$$(3.1) \quad g_q(w, z) = \{q \in N : f(qw) = z\} = \{q \in N : qw \in [z - 1/2, z + 1/2]\} = \left\{q \in N : q \in \left[\frac{z - 1/2}{w}, \frac{z + 1/2}{w}\right]\right\}$$

where  $N = 1, 2, \dots$ . Notice that the length of the interval is  $1/w$ . If  $w > 1$  then the interval is shorter than 1 and only one integer can be contained within. When  $w > 1$  there is no problem since the true value of  $q$  can be recovered exactly. Also, as  $w$  turns to 0, more and more possible values within  $g(w, z)$  are consistent with  $(w, z)$ , hence  $q$  is more imprecisely determined, because  $1/w \rightarrow \infty$  when  $w \rightarrow 0$  continuously. Luckily, there are

few observations with corresponding weight with values below 0.5.

### Illustration

In the GfK dataset, quantities and values for each buy are given after they have been subjected to weighting and rounding. Thus, if an individual buys 10 apples and pays 1695 øre for this and the value of the weighting variable is  $w=0.5$ , then for this observation,  $i$ , the following are given:

$obs^i = (w, z_1 = f(qw), z_2 = f(vw)) = (0.5, f(10*0.5), f(1695*0.5)) = (0.5, 5, 848)$ . The values

$(q, p, v)$  are unknown but  $q$  is according to (3.1) known to lie within:  $q \in \left( \frac{5-1/2}{0.5}; \frac{5+1/2}{0.5} \right) = 9; 11$ . Since

it is also known that  $q$  is an integer,  $q \in 9, 10, 11$ . The interval that contains the true value of  $v$  is determined by the correspondence given in (3.2):

$$(3.2) \quad g_v(w, z_2) = v \in N : f(vw) = z_2 = v \in N : vw \in [z_2 - 1/2, z_2 + 1/2] = \left\{ v \in N : v \in \left[ \frac{z_2 - 1/2}{w}, \frac{z_2 + 1/2}{w} \right] \right\}$$

This implies that  $v$  is known to lie within  $v \in \left( \frac{848-1/2}{0.5}, \frac{848+1/2}{0.5} \right) = 1695, 1697$ . Because  $v$  is an integer,

$v \in 1695, 1696, 1697$ . The interval where price is contained within can also be determined. From the relationship  $v=q*p$ ,  $p$  is contained in the interval given in (3.3):

$$(3.3) \quad g(w, q, z_2) = p \in R_{++} : f(qpw) = z_2 = z_2 - 1/2 \leq qpw < z_2 + 1/2 = \left\{ \frac{z_2 - 1/2}{qw} \leq p < \frac{z_2 + 1/2}{qw} \right\}$$

The price is then known to lie within:

$$p \in \left( \frac{848-1/2}{9*0.5}; \frac{848+1/2}{9*0.5} \right) = 188, \bar{3}; 188, \bar{5} \quad \vee \quad p \in \left( \frac{848-1/2}{10*0.5}; \frac{848+1/2}{10*0.5} \right) = 169,5; 169,7 \quad \vee$$

$$p \in \left( \frac{848-1/2}{11*0.5}; \frac{848+1/2}{11*0.5} \right) = 154,09; 154,27.$$

Since  $q$  and  $v$  can only take 3 different values each,  $p$  must be one of the 9 values in the interior of the following table:

**Table 1** Possible values for observation  $i$ .

	1695	1696	1697
9	188,3333	188,4444	188,5556
10	169,5	169,6	169,7
11	154,0909	154,1818	154,2727

From this example, it is clear that  $p$  does not have to be an integer. When the store manager chooses a pricing policy for a product containing more than one unit, he has two possibilities. He can set a price,  $p$ , for one element contained in the set (product). This will imply that the consumer has to pay  $v=q*p$ , and the price is then an integer. Instead, he can choose, to set the value,  $v$ , which the consumers are required to pay, as is the case in the example. Then the price of one element is  $p=v/q$  and is then mostly not an integer.

An important observation is that if the price is only known to lie in the interval given in (3.3) then the uncertainty of quantity and value cannot be reduced through  $v=p*q$ . On the other hand, if it is known that price cannot be some of the values, then the uncertainty can be reduced and it is often possible to identify the true value of  $q$  and  $v$ .

### Simple approach to reduce uncertainty

To reduce the uncertainty of  $q$  and  $v$  it is necessary to identify the price more exactly than what can be obtained from (3.3) alone. Three possible solutions are:

1. Assume the existence of an observation  $(v_i, q_i)$  where the value and quantity is known exactly and identify the exact price,  $p_i = v_i / q_i$ , for this observation. Other purchases  $q_j \in g_q, v_j \in g_v$  where the true value and quantity is uncertain, but are bought in the same shop the same day as observation  $i$  must have the same price,  $p_i$ . This price can then be used to identify the true quantity and value of observation  $j$  or at least reduce the uncertainty.
2. Assume that prices can only contain integer values.
3. Assume that prices nearest to some reasonable price are most likely.

The first principle can, of course, also be used more generally. As an example, consider five observations which are uncertainly determined and each observation represents purchases in the same shop at the same time; then by using (3.3), each observation gives bounds within which the true price has to be located. The intersection of these bounds gives the tightest possible bound for the price to exist in (given the available information) and if this tightest bound is a subset of any of the original bounds, then we have reduced the uncertainty<sup>10</sup>. This principle uses information from the other observation, which also is the case for the methods discussed in section 4, whereas the last two solutions use information only from the observation itself.

The assumption in solution 2 is obviously not true for all observations, as was seen in the illustration. But, if it is known that prices can only take integer values, this can be used to reduce the uncertainty, or even identify the quantity and the value of a purchase. When a good contains more than one element, the price is not always an integer, whereas goods that only contain one element always have integer prices<sup>11</sup>. Thus, apples - which are often sold in packs of ten or twelve - will not always have integer prices. Watermelons, coconuts and a lot of types of fish are, on the other hand, not sold as often in packs.

The assumption in solution 3 exploits the dramatic price changes that occur when fixing  $v$  and changing  $q$ ; for example a purchase of one litre of milk for the price of 6 kr seems more likely compared to a purchase consisting of two litres of milk with a unit price of 3. This is also why goods bought in small quantities give the largest bias when using the naïve estimator; at the same time it is easier to recover the true values. The opposite is true for quantities bought in larger numbers. Thus, most is gained by abandoning the naïve estimator when goods are

<sup>10</sup> It is the maximum likelihood principle which lies behind, as will be explained in section 4. One problem with this approach is that the weight is the same for all observations bought in the same month, but it can still work because the uncertainty depends not only on the weight, but also on the quantity bought. Also, even if the goods are not bought at the same time, unreasonable prices can still serve as a discriminating device.

<sup>11</sup> An exception is quantity discounts of the type 3 litre of milk for the price  $v$ .

bought in small quantities.

When applying a combination of the approaches in solution 2 and 3 together with (3.1)-(3.3), the true quantity was found in approximately 95% of the observations. One realises that this simple approach was worthwhile when comparing this with the left panel of figure 1.

Obviously, it is possible to apply formula (3.1) to any good not just milk. This was done to recover the possible integer quantities of apples and approximately 46% were uncertainly determined with a few percentage having three possible values and more than 40% having two possible values.

#### 4. An EM algorithm for estimating the probability distributions of observations

One way to interpret the problem faced is that we want to estimate the probability distributions for each observation where the values in the considered distribution are determined in the same way as was done when creating table 1, i.e. by applying equations (3.1)-(3.3). Then one can select the mode value, or the expected value, from these distributions as the most likely true values. But, how do we calculate the probability functions? If we assume that each observation is i.i.d<sup>12</sup> and we had estimates of an overall cumulative distribution function, then the problem would be easy; a probability function for each observation could be calculated as truncated sub-distributions of the cdf. On the other hand, if we had estimated the probability distributions for each observation, then it would be easy to obtain the cdf. This suggests that an approach for estimation is to iterate between calculation of the cdf and the probability functions of each observation. This is the estimation technique chosen in Turnbull (1976) where the main goal is estimation of the cdf and each observation are line segments. Thus, the idea is simply to let each observation consist of a set of indicator variables. For example, take the observation type represented by table 1, this can equivalently be represented by the set:  $obs^i = I_9^{188,3}, I_9^{188,4}, I_9^{188,5}, I_{10}^{169,5}, I_{10}^{169,6}, I_{10}^{169,7}, I_{11}^{154,0}, I_{11}^{154,18}, I_{11}^{154,27}$ , where  $I_9^{188,3} = 1$  if the true quantity is 9 and the true unit price is 188,3 and 0 otherwise, etc. The values of the indicator functions are obviously unknown and the task is to estimate:

\_\_\_\_\_

and this for all observations. An estimator of this is:

$$(4.1) \quad \hat{s}_k = \frac{\hat{s}_k}{\sum_{j \in obs^i} \hat{s}_j}, \forall k \in obs^i, \text{ where on the left side of the equation we have our estimates of the probability}$$

function of the observation (the conditional estimates) and on the right hand side of the equation the estimates of the overall distribution function (the unconditional estimates). The overall distribution function is estimated by:

$$(4.2) \quad \hat{s}_k = \frac{1}{N} \sum_{i=1}^N \frac{\hat{s}_k I_{k \in obs^i}}{\sum_{j \in obs^i} \hat{s}_j}, \forall k \in J = \bigcup_{i=1}^N obs^i, \text{ where } N \text{ is the number of observations and } I_{\cdot} \text{ is an}$$

indicator function. On the right hand side, we have the estimates from equation (4.1). Estimation consists of repeatedly

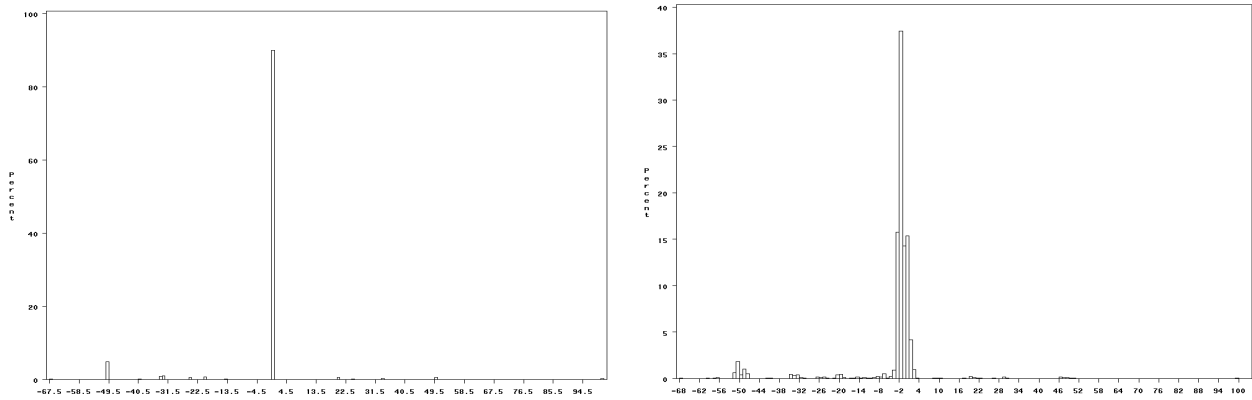
<sup>12</sup> The identically distributed assumption could be relaxed and we might get better results if we split the sample into sub-samples according to some explainable variables that are expected to influence the probability functions; then one could estimate on each of these.

plugging the left hand side value of (4.2) into the right hand side, thereby creating improved estimates; Turnbull (1976) using a Taylor expansion to show that the successive steps give non-decreasing values of a likelihood function defined in the next paragraph. The fix point found by this approach is a consistent estimator of the cdf (Yu (2000)) and we can use this fix point in equation (4.1) to get consistent estimates of the probability distributions of each observation type which are our object of interest.

The likelihood function mentioned which is maximised by successively applying (4.2) is:

## Results using EM algorithm

**Figure 2** Distribution of of all observations based on estimated probability distribution of each observation using EM algorithm. Left using mode value. Right using expected value.



Comparing the distributions in figure 2 with the distribution in the right panel of figure 1, it is clear that the developed methodology of this section gives better results than the naïve estimator.

## 5. Evaluating a proposed EM algorithm

One potential problem with the approach in section 4 is that the probability distribution might be non-unique for some observations. This happens if a set of indicator variables is always contained in the same observations, e.g.

, then it is not possible to estimate  $P I_9^{188,3} = 1$  or  $P I_9^{188,4} = 1$ , but only

$P I_9^{188,3} = 1 \cup I_9^{188,4} = 1 = P I_9^{188,3} = 1 + P I_9^{188,4} = 1$ . This is true because the likelihood function in (4.3) in

this case does not change when changing the estimate of or the estimate of as long as the estimate of

is kept constant.

Li (1997) claims to have found an EM estimator which can solve this problem. Unfortunately this is not the case. As shall be seen, the estimator of Li (1997) is just a sum of the Turnbull (1976) estimator with an additional assumption of

the behaviour of the cdf in the non-unique sets. In fact, the estimator of Li (1997) can be obtained with much less computational resources by simply estimating the Turnbull (1976) estimator and using the mentioned assumption.

We want to estimate the  $P[X^i \leq x]$  and this should ideally be possible for any value  $x \in [0, \infty]$ ; Non-Parametric estimation only allows estimation in the support of data, but we still want to be able to estimate any parameter in a continuous range in the area of support of the data. In practice, we only estimate a finite number of parameters. Thus, the support<sup>13</sup> is,  $x \in [0, 1, \dots, c-1]$ ; since we want to be able to choose any support point in a continuous interval, the support points can be exchanged with any of our choosing. Corresponding to these support points, we define the stochastic (indicator) variables:

$$I_0, I_1, \dots, I_{c-1}, I_k = \begin{cases} 1 & \text{if } X^i \leq k \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad I_0^p, I_1^p, \dots, I_{c-1}^p, I_k^p = \begin{cases} 1 & \text{if } k-1 < X^i \leq k \\ 0 & \text{otherwise} \end{cases}, k = 0, \dots, c-1$$

The probability distribution of observation  $i$  can then be written:

$$(5.1) \quad \Pr[X^i \leq x | l_i, r_i] = \begin{cases} 0, & \text{if } x \leq l_i \\ \frac{\Pr[I_x = 1] - \Pr[I_{l_i} = 1]}{\Pr[I_{r_i} = 1] - \Pr[I_{l_i} = 1]}, & \text{if } x \in [l_i, r_i] \\ 1, & \text{if } x \in [r_i, \infty] \end{cases} = \begin{cases} 0, & \text{if } x \leq l_i \\ \frac{\sum_{j=l_i+1}^x \Pr[I_j^p = 1]}{\sum_{j=l_i+1}^{r_i} \Pr[I_j^p = 1]}, & \text{if } x \in [l_i, r_i] \\ 1, & \text{if } x \in [r_i, \infty] \end{cases}$$

Using the law of iterated expectation:

$$(5.2) \quad \Pr[X^i \leq x] = E[I_x = 1] = E_i[E[I_x = 1 | l_i, r_i]] = E_i[\Pr[X^i \leq x | l_i, r_i]]$$

It is possible to recover the unconditional distribution function. Using the analogous principle we obtain the self-consistent estimator defined in Li (1997).

$$(5.3) \quad \hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{F}(x) - \hat{F}(l_i)}{\hat{F}(r_i) - \hat{F}(l_i)} I_{x \in [l_i, r_i]} + I_{x \in [r_i, \infty]} \right)$$

It is now shown that (5.3) is a sum of Turnbull (1976) estimates. First, it is noticed that the last part of (5.1) is just a sum of conditional probabilities, as given in Turnbull (1976), i.e. the expression is equivalent to:

$$= \left( \frac{s_{l_i+1}}{\sum_{j=l_i+1}^{r_i} s_j} + \dots + \frac{s_x}{\sum_{j=l_i+1}^{r_i} s_j} \right) I_{x \in [l_i, r_i]} + I_{x \in [r_i, \infty]} = \left( \frac{s_{l_i+1}}{\sum_{j=l_i+1}^{r_i} s_j} + \dots + \frac{s_x}{\sum_{j=l_i+1}^{r_i} s_j} \right) I_{x \in [l_i, r_i]} + \frac{\sum_{j=l_i+1}^{r_i} s_j}{\sum_{j=l_i+1}^{r_i} s_j} I_{x \in [r_i, \infty]}$$

Where we have used the definition  $s_x \equiv \Pr[I_x^p = 1]$ , thus (5.3) can be written:

$$\begin{aligned} \hat{P}r[X^i \leq x] &= \frac{1}{n} \sum_{i=1, \dots, n} \left( \frac{s_{l_i+1}}{\sum_{j=l_i+1}^{r_i} s_j} + \dots + \frac{s_x}{\sum_{j=l_i+1}^{r_i} s_j} \right) I_{x \in [l_i, r_i]} + \frac{1}{n} \sum_{i=1, \dots, n} \frac{\sum_{j=l_i+1}^{r_i} s_j}{\sum_{j=l_i+1}^{r_i} s_j} I_{x \in [r_i, \infty]} \\ &= \frac{1}{n} \sum_{i=1, \dots, n} \frac{s_0 I_{0 \in [l_i, r_i]}}{\sum_{j=l_i+1}^{r_i} s_j} + \dots + \frac{1}{n} \sum_{i=1, \dots, n} \frac{s_x I_{x \in [l_i, r_i]}}{\sum_{j=l_i+1}^{r_i} s_j} = \hat{s}_0 + \dots + \hat{s}_x \end{aligned}$$

<sup>13</sup> For convenience, we consider the univariate case.

The last equation follows from the self-consistency of the Turnbull (1976) estimator, i.e.

$$(5.4) \quad \hat{s}_k = \frac{1}{n} \sum_{i=1, \dots, n} \frac{\hat{s}_k I_{x \in l_i, r_i}}{\sum_{j=l_i+1}^{r_i} \hat{s}_j}, k = 0, \dots, c-1$$

Which follows from the law of iterated expectation:

$$(5.5) \quad \Pr[x-1 < X^i \leq x] = E[I_x^p = 1] = E_i[E[I_x^p = 1 | l_i, r_i]] = E_i[\Pr[x-1 < X^i \leq x | l_i, r_i]]$$

Noting that:

$$\Pr[X^i \leq x] = \Pr\left[\bigcup_{i=1}^x (i-1 < X^i \leq i)\right] = \sum_{i=1}^x \Pr[i-1 < X^i \leq i] = s_0 + \dots + s_x$$

And observing that the estimator in (5.4) is based on (5.5) and the estimator in (5.3) is based on (5.2) it is not surprising that the estimator of Li (1997) is just the sum of Turnbull(1976) estimates:

$$\begin{aligned} \Pr[X^i \leq x] &= E_i[\Pr[X^i \leq x | l_i, r_i]] = E_i\left[\sum_{j=1}^x \Pr[j-1 < X^i \leq j | l_i, r_i]\right] \\ &= \sum_{j=1}^x E_i[\Pr[j-1 < X^i \leq j | l_i, r_i]] = \sum_{j=1}^x E_i[\Pr[I_j^p = 1 | l_i, r_i]] = s_0 + \dots + s_x \end{aligned}$$

The reason why the approach in Li (1997) is a sum of Turnbull (1976) estimates is that although the initial variable considered is continuous, the uncertainty of each observation transforms its distribution into a multinomial distribution in both approaches.

## 6. Conclusion

This paper dealt with a problem in the data delivered by the **market research** company Gfk Denmark. An analysis of a naïve estimator formerly used and a new methodology based on the Expectation Maximisation (EM) algorithm revealed that the new estimation technique is superior.

The paper has shown how to calculate the probability distribution of each observation when the exact value of each observation might be uncertain. The probability distribution of each observation is calculated using an EM algorithm, exploiting the fact that each observation comes from the same data generating process, which means that the distribution of each observation is a sub-distribution of the overall data generating distribution.

## References

- Dempster A. P.; Laird N. M.; Rubin D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1. (1977), pp. 1-38.
- EFRON B. . (1967). The two-sample problem with censored data. Proc. 5th Berkeley Symposium on Math.Statist. and Prob., 4, 831-853.
- EFRON, B. (1982): "Maximum Likelihood and Decision Theory," Annals of Statistics, 10, 323-339.
- Kaplan E. L. and Meier P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American

Statistical Association, Vol. 53, No. 282 (Jun., 1958), pp. 457- 481

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of Royal Statistical Society.*, 38, 290-295.

Li, L., Watkins, T. and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics.* 24, 531-542.

Wu C. F. Jeff (1983). On the convergence properties of the EM algorithm. *The annals of statistics*, Vol. 11, No 1 (Mar., 1983), 95-103.

**Yu, Qiqing; Li, Linxiong; Wong, George.** (2000). On consistency of the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics.* 27, 35-44.



# Paper 3: Estimation of multi-dimensional discrete stochastic variables

---

Aslak Hedemann Hansen  
University of Copenhagen

## **Abstract:**

This paper deals with the estimation of multi-dimensional discrete stochastic variables. It is well known that when the number of observations is small relative to the number of parameters that are to be estimated, it can be an advantage in finite sample applications to use a method that smoothes parameter estimates using a kernel function. Least square cross-validation is an example of such a smoothing approach which leads to differing asymptotic behaviour of the smoothing parameters depending on whether each variable is uniformly distributed or not. Also, smoothing parameters, which correspond to uniformly distributed variables, have positive probability of not converging to their optimal value when using least square cross-validation. This paper shows why this is the case and derives the asymptotic probability distribution of the smoothing parameter. Furthermore, a criterion function is suggested that solves this problem. The fact that the smoothing parameters now asymptotically converge to the optimal values effectively means that the associated estimator shrinks to the smallest variance unbiased frequency estimator and lends support to the usability of smoothing methods in discrete variable applications.

## 1. Introduction

The probability distribution of a multidimensional discrete stochastic variable can be estimated by the maximum likelihood estimator (MLE), also known as the frequency estimator, since the probability of observing a given value in the distribution is estimated by the frequency of observations with this value. It is well known that when the number of observations is small relative to the number of parameters to estimate, as is often the case in economics applications, it can be an advantage in finite sample applications to use a method that smoothes parameter estimates using a kernel function. This is further supported by the fact that the frequency estimator asymptotically has the same or lower efficiency, depending on the probability distribution, than certain smoothing estimators. For example, for certain probability distributions, such as when one of the variables is uniformly distributed (a definition is given in expression (2.1)), smoothing-based unbiased estimators with smaller variance do exist.

This paper analyses such a smoothing-based estimator that is identical to the least square cross-validation estimator seen in Ouyang *et al.* (2006). Least square cross-validation leads to differing asymptotic behaviour of the smoothing parameters depending on whether each variable is uniformly distributed or not. The reason for this is that, in the population problem, the values of the smoothing parameters that minimise the integrated mean square error of the estimator, the optimal values, are zero for smoothing parameters which correspond to non-uniformly distributed variables when the number of observations goes to infinity, and always one for smoothing parameters which correspond to uniformly distributed variables. The reason for this is that the optimal smoothing parameters balance the trade-off between smaller variance and larger biases that are the result of increasing a smoothing parameter and since smoothing parameters, which correspond to uniformly distributed variables, are not related to the bias term, there is no trade-off and their value should be the maximum one, i.e. minimise variance. Smoothing parameters which correspond to non-uniformly distributed variables, on the other hand, are related to the bias term and since the variance term goes to zero as the number of observations goes to infinity, these parameters should be set at zero in this case. When moving to the actual estimation of a sample, it is natural to follow the same approach as in the population problem and to simply substitute the function that was minimised in the population problem with an estimator hereof. This is, as we will see, the same as the approach of Ouyang *et al.* (2006). Varying asymptotic behaviour of the smoothing parameters depending on whether a variable is uniformly distributed or not is therefore expected. Unfortunately, as shown in Ouyang *et al.* (2006), the smoothing parameters which correspond to uniformly distributed variables, do not always converge to their optimal value. The purpose of this paper is to derive the probability distribution of smoothing parameters which correspond to uniformly distributed variables chosen by least square cross validation and to show why they do not converge to their optimal value. Furthermore, a new criterion function is introduced that leads the smoothing parameters to asymptotically converge to their optimal values and effectively means that the associated estimator shrinks to the smallest variance unbiased frequency estimator, thereby lending support to the usability of smoothing methods in discrete variable applications.

## 2. Estimation of multi-dimensional discrete stochastic variables

### 2.1 Notation and statistical model

We have,  $\mathbf{X}$ ,  $r$ -dimensional discrete stochastic variables,  $\mathbf{X} \in \mathcal{X}$ .  $\mathbf{X}$  has support  $\mathcal{X}$ ,

where  $\mathbf{X}$ . It is assumed that the stochastic variables are independent and identically distributed, and that a subset of  $\mathbf{X}$  is uniformly distributed. A variable  $X_j$  is uniformly distributed with respect to  $\mathcal{X}_j$ , if and only if :

---

A subset of the variable  $X_j$  is indicated by  $\mathcal{X}_j$ , here the set  $\mathcal{X}_j$  can be written  $\mathcal{X}_j = \{x_j \in \mathcal{X}_j\}$ , where  $\mathcal{X}_j$  is the set of uniformly distributed variables. Note that  $\mathcal{X}_j$  can be composed in  $r$  different ways, since there are  $r$  variables, and each of them can be uniformly distributed. Closely related to the problem of finding an efficient estimator is the problem of finding the variables in the multi-dimensional distribution that are uniformly distributed. If it is, for instance, known that  $\mathcal{X}_j$ , the estimator:

can be used, where  $\mathbf{X}$ , and  $\mathbf{I}$  is the indicator function that equals one if the argument is true and zero otherwise.  $\hat{\theta}_{ML}$  has the same bias, if the correct  $\mathcal{X}_j$  is applied, but smaller variance than the frequency estimator (MLE),  $\hat{\theta}_{ML}$ . This is clear, since:

---

Whereas the frequency estimator,  $\hat{\theta}_{ML}$ , converges to a normal distribution with the same mean, but, with greater variance  $\sigma^2$ . A more efficient estimator can then be determined when  $\mathcal{X}_j$  is determined. If  $\mathcal{X}_j$  is, on the contrary, unknown, it is possible to define  $r$  potential estimators from (2.2) where only one fully exploits that a subset of the variables are uniformly distributed. To solve this problem, a smoothing parameter is defined for each of the  $r$  potential estimators,  $\lambda_j$ . A procedure that asymptotically finds the correct hypothesis, which variables that belong to  $\mathcal{X}_j$ , is a procedure that asymptotically will let  $\lambda_j$  converge to one, for  $\mathcal{X}_j$ , and the other  $r-1$  smoothing parameters,  $\lambda_{j'} \rightarrow 0$ , converge to zero. To this end, the following kernel function is defined:

with the corresponding probability function:

where  $\mathbf{X}$  og  $\mathbf{X}$  is the frequency estimator for estimation of the probability  $\theta_j$ . With convention  $\mathbf{X} \cap \mathbf{X} = \mathbf{X}$  and  $\mathbf{X} \cup \mathbf{X} = \mathbf{X}$ . Here  $\mathbf{X}$  and  $\mathbf{X}$  is the power set of  $\mathcal{X}_j$ , i.e.  $\mathcal{X}_j$ .

A commonly used kernel function is the product kernel function used in, e.g. Hall *et al.* (2004) and Ouyang *et al.* (2006). This kernel function can easily be derived from (2.4) by using expression (2.6) below. Let  $w_j$  be the weight (probability) that is put on the hypothesis that  $\mathcal{X}_j$ . Consider the following expression for  $\hat{\theta}_j$ ,

Let  $w_j$  be the weight (probability) put on the hypothesis that  $\theta_j$  is uniformly distributed ( $\theta_j \sim U(\theta_{j1}, \theta_{j2})$ ). From (2.4), (2.6) and (2.7) it is clear that:

With convention  $\theta_j = \theta_{j1}$  and  $\theta_j = \theta_{j2}$ . By using expression (2.6), it is possible to substitute in (2.4) with  $\theta_j$ , and (2.4) is reduced to the kernel function:

with the corresponding probability function:

Again  $\hat{w}_j$  is the frequency estimator for estimation of the probability  $w_j$  and with convention  $\theta_j = \theta_{j1}$  and  $\theta_j = \theta_{j2}$  is the power set of  $\theta_j$ . Note that the kernel function defined in (2.8) and (2.7) can also be written as  $K(\theta_j, \theta_j)$ , and with substitution  $\theta_j = \theta_{j1}$ , it is easy to show that:

$$K(\theta_j, \theta_j) = \frac{1}{\theta_{j2} - \theta_{j1}} \int_{\theta_{j1}}^{\theta_{j2}} \frac{1}{\theta_j} d\theta_j$$

The conclusion is that the kernel function in (2.8) is the same as the one used in Hall *et al.* (2004) and Ouyang *et al.* (2006) for the smoothening of multi-dimensional discrete variables. A procedure that asymptotically finds the correct hypothesis is a procedure that asymptotically will let the smoothening parameters which correspond to uniformly distributed variables converge to one and the others converge to zero.

## 2.2 Extending the model to include marginally uniformly distributed variables

As mentioned in the introduction and as will be explored in more detail in section 3.1, least square cross validation does not always asymptotically lead to the correct hypothesis as there is positive probability that smoothening parameters which correspond to uniformly distributed variables do not converge to one. In section 3.1, how to choose a criterion function, such that the correct hypothesis as defined in section 2.1 is asymptotically recovered, will also be discussed. Going back to the introduction of the statistical model introduced in section 2.1, it is clear that it does not distinguish between cases where a variable,  $\theta_j$ , is not uniformly distributed, but is marginally uniformly distributed and in cases where a variable,  $\theta_j$ , is not even marginally uniformly distributed. To be clearer, assume that  $\theta_j$  does not satisfy (2.1), i.e.  $\theta_j$  is not uniformly distributed (with respect to  $\theta_j$ ). However,  $\theta_j$  does satisfy (2.1) and:

clearly,  $\theta_j$  is uniformly distributed with respect to the set  $\theta_j$ . In this case, a procedure that identifies the correct hypothesis, as defined in section 2.1, will let  $w_j$  converge to zero and this would also be the case for a different  $r$ -

dimensional probability distribution where only satisfies \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_ or a  $r$ -dimensional probability distribution where does not even satisfy \_\_\_\_\_. However, these distinct cases can be identified by using a procedure that identifies the correct hypothesis, as defined in section 2.1, but on a subset of the variables. To be concrete and see the value of including marginally uniformly distributed variables in the model, assume that we have a procedure, \_\_\_\_\_, that can take a set of iid variables from the distribution \_\_\_\_\_, and identify which of the \_\_\_\_\_ are uniformly distributed with respect to \_\_\_\_\_, and which are not. As already mentioned, such a procedure that does this asymptotically will be given in section 3.1.

If interest is on a  $r$ -dimensional probability distribution, it could be useful to know if any of the variables are uniformly distributed and if any of the non-uniformly distributed variables are marginally uniform. The following simple procedure provides the answer to this question. Let  $\mathbf{K}$  be the set of variables that are uniformly distributed with respect to the set  $\mathbf{R}$ , and  $\mathbf{V}$  be the set of variables that are not uniformly distributed with respect to the set  $\mathbf{R}$ , but is marginally uniformly distributed. Define:

Initial set \_\_\_\_\_, \_\_\_\_\_.

Step 1: \_\_\_\_\_;

For each \_\_\_\_\_ do: If \_\_\_\_\_ then add  $l$  to \_\_\_\_\_; end do;

If \_\_\_\_\_, then stop, else proceed to step 2.

Step 2: For each \_\_\_\_\_ do:

\_\_\_\_\_; If \_\_\_\_\_ then add  $l$  to \_\_\_\_\_; end do;

If \_\_\_\_\_, but \_\_\_\_\_, e.g. \_\_\_\_\_, then the estimator \_\_\_\_\_ is more efficient than the frequency

estimator, in general if \_\_\_\_\_, \_\_\_\_\_, the following estimator could be used \_\_\_\_\_.

Note that the estimator exploits the fact that it is known that a subset of the variables are marginally uniformly distributed, which ensures that the estimates are in fact marginally uniform and decreases variability. If \_\_\_\_\_ then the estimator is not very useful, thus the estimator can be slightly modified for this reason to:

---

which is more efficient than the estimator \_\_\_\_\_ defined in \_\_\_\_\_ if \_\_\_\_\_ or \_\_\_\_\_ or \_\_\_\_\_. To see that this described procedure can improve the estimation efficiency relative to the procedure described in section 2.1, some examples will be considered where the estimator \_\_\_\_\_ defined in (2.12) is an improvement on the estimator \_\_\_\_\_ defined in \_\_\_\_\_.

For instance, assume that a probability distribution contains no uniform distributed variables, but does contain variables that are uniformly distributed with respect to a subset of the variables. Then the procedure in section 2.1 will simply turn up with the frequency estimator and no efficiency gains and since there exist variables that are uniformly distributed with respect to a subset of the variables, it would have been possible to use an estimator, \_\_\_\_\_, as defined in (2.12), which would produce efficiency gains relative to the frequency estimator. For another example where the procedure of section 2.1 does in fact produce efficiency gains relative to the frequency estimator, but where even more

efficiency can be obtained, consider a probability distribution where a subset of the variables are uniformly distributed, and a subset of the non-uniformly distributed variables are marginally uniformly distributed. In this case, the procedure of section (2.1) and the associated estimator given in formula (2.2) will produce efficiency gains relative to the frequency estimator, but they do not fully exploit the fact that some of the non-uniformly distributed variables are marginally uniform and an estimator such as  $\hat{\theta}_{\text{MUE}}$ , will produce efficiency gains relative to the estimator  $\hat{\theta}_{\text{freq}}$ . It will now be shown through Monte Carlo simulations that the procedure described in this section and the associated estimator  $\hat{\theta}_{\text{MUE}}$  can create efficiency gains when a subset of the variables is marginally uniformly distributed.

### Monte Carlo simulations

In this section, we report some simulations designed to examine the finite sample performance of 3 estimators:

I will first consider the following cases:

Case (i): A simple case which has one non-uniform variable and one uniform variable, both of which are binary:

with  $X_1 \sim \text{Bernoulli}(0.5)$  and  $X_2 \sim \text{Bernoulli}(0.5)$ .

Case (ii): in this case, there are no uniformly distributed variables, but  $X_1$  is marginally uniform:

$X_1 \sim \text{Bernoulli}(0.5)$ ,  $X_2 \sim \text{Bernoulli}(0.5)$ ,  $X_3 \sim \text{Bernoulli}(0.5)$ ,  $X_4 \sim \text{Bernoulli}(0.5)$ ,  $X_5 \sim \text{Bernoulli}(0.5)$ ,  $X_6 \sim \text{Bernoulli}(0.5)$ ,  $X_7 \sim \text{Bernoulli}(0.5)$ ,  $X_8 \sim \text{Bernoulli}(0.5)$ ,  $X_9 \sim \text{Bernoulli}(0.5)$ ,  $X_{10} \sim \text{Bernoulli}(0.5)$ .

The number of simulations is  $M=1000$  and the sample sizes are  $n=50, 100, 200, 500$  and  $1000$ . I compute the average mean squared error (MSE) by  $\frac{1}{M} \sum_{j=1}^M \text{MSE}_j$ , where  $\text{MSE}_j$  is the estimated MSE in the  $j$ th simulation.

Table 2.1. MSE summary for case (i).

$n$	$\hat{\theta}_{\text{freq}}$	$\hat{\theta}_{\text{MUE}}$	$\hat{\theta}_{\text{MUE}}^*$
50	0.01478	0.5667533	0.2833829
100	0.0070958	0.6138163	0.2885933
200	0.00343405	0.6326996	0.3008401
500	0.001437512	0.5800173	0.2903684
1000	0.000703264	0.5984496	0.3043949

Table 2.2. MSE summary for case (ii).

$n$	$\hat{\theta}_{\text{freq}}$	$\hat{\theta}_{\text{MUE}}$	$\hat{\theta}_{\text{MUE}}^*$
50	0.0139392	0.5898609	3.222308
100	0.0073248	0.5850221	5.797919
200	0.00339325	0.591459	12.14147
500	0.00139168	0.606639	29.08334
1000	0.000691368	0.5673642	58.18888

It is clear from table 2.1 and 2.2 that the estimator that exploits that  $X_1$  is marginally uniform,  $\hat{\theta}_{\text{MUE}}$ , and has a better finite sample performance than the frequency estimator.

In case 1, where  $X_1$  is not only marginally uniform, but also uniformly distributed, i.e.  $X_1 \sim \text{Uniform}(0,1)$  is uniformly distributed with

respect to  $\hat{\theta}_1$ , the finite sample performance of the estimator that is especially designed for this case,  $\hat{\theta}_1$ , is approximately twice as good as the estimator that only exploits the fact that  $\theta$  is marginally uniform,  $\hat{\theta}_2$ . In case 2, it is also clear that the estimator  $\hat{\theta}_2$  has poor performance and this is because the estimator is biased in this case.

I will now consider two simple cases with one non-uniform variable and one uniform variable of varying degrees of the support of the distribution:

Case (i.1): the non-uniformly distributed variable is binary and the support of the uniform distribution is  $[0, 1]$  or  $[0, 0.5]$ . The probability distribution is given by:  $p_1 = 0.5$  with  $p_2 = 0.5$ .

Case (i.2): the non-uniformly distributed variable has support  $[0, 0.5]$  and the support of the uniformly distributed variable is  $[0, 1]$  or  $[0, 0.5]$ . The probability distribution is given by:  $p_1 = 0.5$  with  $p_2 = 0.5$ ,  $p_3 = 0.5$ , and  $p_4 = 0.5$ .

Table 2.3. MSE summary for case (i.1)-(i.2).  $n=500$ .

(c1,c2)			
Case (i.1): (2,2)	0.001437512	0.5800173	0.2903684
Case (i.1): (2,4)	0.001671072	0.5062134	0.1247846
Case (i.1): (2,32)	0.001967456	0.4592336	0.01381378
Case (i.2) (4,2)	0.001693272	0.836663	0.4127039
Case (i.2) (4,4)	0.001854504	0.759564	0.1841096
Case (i.2) (4,32)	0.001982245	0.7583532	0.02302061

Once again, it is clear that  $\hat{\theta}_1$  dominates the other estimators when one of the variables is uniformly distributed. I will now consider four simple cases with varying degrees of support and which have two non-uniformly distributed variables, but with one being marginally uniform:

Case (ii.1): in this case,  $\theta$  is marginally uniform and the probability distribution is:  $p_1 = 0.5$ ,  $p_2 = 0.5$ ,  $p_3 = 0.5$ , and  $p_4 = 0.5$ .

Case (ii.2): in this case  $\theta$  is marginally uniform with probability distribution:  $p_1 = 0.5$ ,  $p_2 = 0.5$ ,  $p_3 = 0.5$ , and  $p_4 = 0.5$ .

Case (ii.3): in this case  $\theta$  is marginally uniform with probability distribution:  $p_1 = 0.5$ ,  $p_2 = 0.5$ ,  $p_3 = 0.5$ , and  $p_4 = 0.5$ .

Case (ii.4): in this case  $\theta$  is marginally uniform and the probability distribution,  $p_1 = 0.5$ ,  $p_2 = 0.5$ ,  $p_3 = 0.5$ , and  $p_4 = 0.5$ , is best summarised in a table:

	.12	.09	.07	.10
	.03	.07	.06	.06
	.08	.06	.06	.05
	.02	.03	.06	.04

Table 2.4. MSE summary for case (ii.1)-(ii.4).  $n=500$ .

(c1,c2)			
(ii.1): (2,2)	0.00139168	0.606639	29.08334
(ii.2): (2,4)	0.001667032	0.4504497	15.13967
(ii.3): (4,2)	0.001727744	0.8247795	7.305764
(ii.4): (4,4)	0.001849008	0.7754922	2.118995

Once again it is clear that  $\hat{\beta}_{OLS}$  dominates the other estimators when one of the variables is marginally uniform and when none of the variables are uniformly distributed.

### 3. Criterion function

Minimising the criterion function:

will lead the smoothening parameters to converge to their optimal values<sup>14</sup>. Least square cross validation consists of minimising the criterion function that arises from (3.1) by substitution of terms of the type

$\hat{\beta}_{OLS}$ , in (3.1) with an unbiased estimator hereof  $\hat{\beta}_{OLS}$ . It is straightforward to show that this is the criterion function used in Ouyang *et al.* (2006). As mentioned in the introduction, there is a positive probability that smoothening parameters which correspond to uniformly distributed variables will not converge to their optimal value using least square cross validation. There is insufficient smoothening in this case and intuitively it might be a good idea to choose a criterion function that puts relatively more (less) weight on variance (bias) minimisation. With this in mind, I introduce criterion function  $C(\hat{\beta})$  and probability estimator  $\hat{p}$

Note that the estimator  $\hat{\beta}_{OLS}$  is defined in (2.9) and is identical to the estimator  $\hat{\beta}_{OLS}$  for  $\hat{\beta}_{OLS}$ . Note that  $C(\hat{\beta})$  is identical to the criterion function used in Ouyang *et al.* (2006) and incidentally minimising this criterion function with respect to  $\hat{\beta}$  is least square cross validation.

### 3.1 Analysis

Expression (A3.6) in the appendix show that smoothening parameters which correspond to non-uniformly distributed variables are of smaller order when  $\hat{\beta}_{OLS}$  and by using this, the following version of the criteria function is derived in lemma A3:

Where  $\hat{\beta}_{OLS}$  is a term of a smaller order (the exact expression of  $\hat{\beta}_{OLS}$  is given in lemma A3 in the appendix) and

$\hat{\beta}_{OLS}$ . The function  $C(\hat{\beta})$  is defined in (A1) of Lemma A1.1,  $\hat{\beta}_{OLS}$  and  $\hat{\beta}_{OLS}$  in Lemma A1.3,  $\hat{\beta}_{OLS}$  and in (A3.4) of Lemma A3.1,  $\hat{\beta}_{OLS}$  in Lemma A4. Before showing that the smoothening parameters that minimises the criterion function  $C(\hat{\beta})$  will converge to their optimal values if the auxiliary parameters  $\hat{\beta}_{OLS}$  are chosen in a certain way, it is now a good opportunity to discuss in more detail why least square cross validation does not lead the smoothening parameters which correspond to uniformly distributed variables to converge to their optimal values. Least square cross validation solves the problem  $\hat{\beta}_{OLS}$ . When  $\hat{\beta}_{OLS}$  expression (A3) changes to:

<sup>14</sup> The bias term and variance term are defined in respectively (A1) in Lemma A1.1 and (A1.5) in Lemma A1.2.



Where

In lemma A1.3 it is proven that  $\hat{\theta}_n$  is only a function of  $\theta$  and that the global minimum point for this function is for  $\theta = \theta^*$ . It is therefore clear from (3.4) that  $\hat{\theta}_n$  is at least of the order  $O_p(n^{-1/2})$ , but the rate could be faster if the stochastic variable  $\varepsilon_t$  was always positive. It can be proven that  $\hat{\theta}_n$  is strictly of the order  $O_p(n^{-1/2})$ . Consequently, the rate of  $\hat{\theta}_n$  is strictly of the order  $O_p(n^{-1/2})$ . In view of the rate of  $\hat{\theta}_n$ , it is clear from <sup>15</sup> that the behaviour of the smoothening parameters which correspond to uniformly distributed variables,  $\hat{\theta}_n$ , are determined as the argument that minimises the third term:

Also,  $\hat{\theta}_n$  (see lemma A3.3), thus if  $\hat{\theta}_n$  is also minimised by  $\theta = \theta^*$ , then  $\hat{\theta}_n$  converges to its optimal values, this has positive probability, but  $\varepsilon_t$  is stochastic and can take values such that the minimisation of (3.5) does not lead  $\hat{\theta}_n$  to converge to its optimal value,  $\theta = \theta^*$ . However,  $\hat{\theta}_n$  would converge to its optimal value if  $\varepsilon_t$  as  $\theta^*$ . Thus, if expression (A3) is true, i.e. if  $\varepsilon_t$  is exchanged with  $\theta^*$  and because  $\varepsilon_t$  as  $\theta^*$  for  $\theta^*$ ,  $\hat{\theta}_n$  would converge to its optimal value.

I now return to the more general problem where  $\varepsilon_t$ .

### Result 3.1

Let  $\varepsilon_t$  be defined as in (3.2). Assume without lose of generality that  $\varepsilon_t$ . Assume that,  $\varepsilon_t$ . Then:

In addition assume that,  $\varepsilon_t$ . Then:

### Proof

$\varepsilon_t$  ensures that (A3) is true, hence:

The order of  $\hat{\theta}_n$  can therefore be determined by the order of  $\varepsilon_t$  and  $\theta^*$ . Lemma A3.7 show that:

<sup>15</sup> The first term is unrelated to  $\theta$  and the second term is of the order  $O_p(n^{-1/2})$  faster than the third term.

Where  $\mathbf{f}$  is a sum of functions of  $\mathbf{x}$  that is of smaller order when minimising  $\mathbf{f}$  and  $\mathbf{g}$  is a cross matrix with entries  $f_{ij}$  if  $i \neq j$  and  $g_{ii} = 1$  if  $i = j$ , and  $\odot$  is the elementwise multiplication operator. Lemma A3.8 show that  $\mathbf{f} = \mathbf{g}$  with equality if and only if  $\mathbf{f} = \mathbf{g}$ . Thus, the order of  $\mathbf{f}$  can be determined by:

Expression (3.9) leads to the first order condition

Where  $\mathbf{f}$ . Letting  $\mathbf{f}$  denote the value of  $\mathbf{f}$  that solves  $\mathbf{f} = \mathbf{g}$ , then  $\mathbf{f}$ , where  $\mathbf{f}$ . Note that  $\mathbf{f}$ , where  $\mathbf{f}$  from Lemma A3.4, thus,  $\mathbf{f}$ . Define  $\mathbf{f}$ , and insert  $\mathbf{f}$  in expression  $\mathbf{f}$  gives  $\mathbf{f}$ , thus,  $\mathbf{f}$ . So:

This proves  $\mathbf{f}$ .  
To prove  $\mathbf{f}$  assume  $\mathbf{f}$ , then:

This shows that terms related to  $\mathbf{f}$  in the sum  $\mathbf{f}$  is of the order  $\mathbf{f}$ , hence (3.8) and (3.12) shows that

Q.E.D.

### Result 3.2

Let  $\mathbf{f}$  be an identical and independent distributed sample. A necessary and sufficient condition for (3.13) to be true is (3.14).

### Proof

Expression (3.15) and (3.16) are proven in Lemma A3.5 and A3.6.

Let  $\{X_i\}_{i=1}^n$  be an identical and independent distributed sample, then

Where  $\hat{\mu}_n$  and  $\hat{\sigma}_n^2$ .

(3.15) and (3.16) shows that  $\hat{\mu}_n$ , also,

, so a necessary condition for  $\hat{\mu}_n$  to be true is  $\hat{\sigma}_n^2 > 0$ . From expression (3.17) it is clear that

reduces to:

Where  $\hat{\mu}_n$ , especially  $\hat{\mu}_n$ , and

from  $\hat{\mu}_n$ , thus,  $\hat{\mu}_n$ .

However,  $\hat{\mu}_n$  and

, hence a necessary condition for (3.13) to be true is  $\hat{\sigma}_n^2 > 0$ . This also shows that (3.14) is a sufficient condition for (3.13) to be true.

Q.E.D.

### Result 3.3

Assume that,  $\hat{\mu}_n$ ,  $\hat{\sigma}_n^2$ , and  $\hat{\mu}_n$ , then

—

### Proof

Result 3.1 show that  $\hat{\mu}_n$  satisfy condition (3.14), which implies that (3.13) is satisfied by result 3.2.

From this it is clear that — —

—

Q.E.D.

Note that result 3.3 shows that if the smoothening parameters are chosen as stated in result 3.3, e.g.

and  $\hat{\mu}_n$ , then the estimator in (2.9) would

converge to a distribution which is asymptotically equivalent to the normal distribution that the estimator in (2.2), with correct  $k=N$ , would also converge to.

### Result 3.4

Assume that  $\hat{\mu}_n$ ,  $\hat{\sigma}_n^2$ . Then:

Where:

is a chi-square distribution with degrees of freedom and:

### Proof

First,

follows directly from result 3.1. To find the probability distribution of remember (3.8) and observe that (3.21) implies that . Thus:

Note that and is unrelated to , thus

Where . Define — then expression (A5.3) in Lemma A5 show that: , so:

Remember the definition of in expression (3.20) and inserting in (3.25) yield:

To find three cases are considered. Case 1: Assume .

Then reduces to . Case 2: Assume and which has already been covered in case 1. Then it is clear that the two functions in the sum in (3.26) are minimised by , so for this case.

Case 3: Assume , then —, follows from (3.27). Expression then follows from (A5.5) and (3.24).

**Q.E.D.**

## Probability distribution of smoothing parameter

In this section the probability distribution of a smoothing parameter  $\hat{h}$  which corresponds to the only uniformly distributed variable with its auxiliary parameter equaling zero  $\hat{h}_0$  is derived for some specific probability distributions using result 3.4. Two cases are considered and both cases considers bivariate distributions with the second variable being uniformly distributed and  $\hat{h}_0$  and  $\hat{h}_1$ .

Case (i): A bivariate distribution which has one non-uniform variable and one uniform variable, both of which are binary:  $\hat{h}_0$  with  $\hat{h}_1$ .

Case (ii): same as case (i) except that  $\hat{h}_0$  is not binary, i.e.  $\hat{h}_0 \in [0, 1]$ , so that  $\hat{h}_1 \in [0, 1]$  with  $\hat{h}_0$ .

In both cases there is only one uniformly distributed variable, and setting  $\hat{h}_0 = \hat{h}_1$ , reduces (3.8) to  $\hat{h} = \hat{h}_0$  and the solution is:

Consequently, the distribution of  $\hat{h}$  is determined when the distribution of  $\hat{h}_0$  is determined. By expression (A5.5) in Lemma A5  $\hat{h}$  converges to a weighted sum of chi-square distributed variables and in both cases:

Where  $\chi^2_r$  is a chi-square distribution with  $r$  degrees of freedom. However,  $\hat{h}$  differs in the two cases and consequently so does the distribution of  $\hat{h}$ . The distribution of  $\hat{h}$  heavily relies on the distribution of a linear combination of chi-square random variables. Ruben (1962) suggested a method to calculate the probability that a linear combination of  $m$  chi-square variables will be smaller than a constant such as  $\hat{h}_0$ . Farebrother (1984) utilizes this method and creates an improved version AS 204 of the algorithm AS106 described by Sheil and O'Muircheartaigh (1977). The algorithm is available at: <http://lib.stat.cmu.edu/apstat/204>. Using this algorithm I find that Pr

$\hat{h} \leq \hat{h}_0$  in case (i) and  $\hat{h} \leq \hat{h}_1$  in case (ii).

Ouyang *et al.* (2006) report some simulations to examine the finite sample performance of the cross validation method and to assess the probability that the cross validation method selects  $\hat{h}_0$  at its maximum value. They consider the same probability distributions as case (i) and (ii) above, hence, when a large number of simulations with large sample sizes are used the simulated probabilities should be close to the calculated probabilities using algorithm AS 204. In fact, Ouyang *et al.* (2006) page 78-79 does  $M=1000$  simulations and finds that the simulated probabilities are 0.641, 0.636 and 0.642 for sample sizes  $n=50, 100$  and  $200$  in case (i) and 0.601, 0.600 and 0.606 for  $n=50, 100$  and  $200$  in case (ii).

If there is only one variable and this variable is uniformly distributed then result 3.4 and expression (A5.4) in Lemma A5 show that if  $\hat{h}_0 = \hat{h}_1$ , then:

$\hat{h} = \hat{h}_0$ , where  $\chi^2_r$  is a chi-square distribution with  $r$  degrees of freedom.

**Table 3.1**  $\hat{h}$  as a function of the number of cells  $\hat{h}_0$ .

	2	3	4	5	6	7	8	9	10	11
	0.6826	0.6321	0.6083	0.5939	0.5841	0.5768	0.5711	0.5665	0.5627	0.5595

	500	5000	100000	10000000000
	0.50842	0.50266	0.50059	0.50000

Table 3.1 and expression  $\hat{h} = \hat{h}_0$  shows that the probability that the smoothing parameter will converge to its optimal value is largest when the number of cells  $\hat{h}_0$  is small and stabilises around  $\frac{1}{2}$  when the size of the support goes to infinity.

### Asymptotical distribution of $\hat{\beta}_n$ using least square cross validation

In result 3.3 the asymptotical distribution of  $\hat{\beta}_n$  was derived for certain values of  $\lambda$ . If the restriction on  $\lambda$  in result 3.3 is not satisfied then the asymptotical distribution of  $\hat{\beta}_n$  will be more complex. In this section the asymptotical distribution of  $\hat{\beta}_n$  is derived for least square cross validation, i.e.  $\lambda = \lambda_{LSCV}$ , for the case  $\lambda \rightarrow 0$ . I will consider a less restrictive case than that will have the same asymptotical distribution as the least square cross validation case when  $\lambda \rightarrow 0$ . Assume that:  $\lambda \rightarrow 0$  and  $\lambda \lambda_n \rightarrow 0$ . Then we know the asymptotic distribution of  $\hat{\beta}_n$  from result 3.4, i.e.  $\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow N(0, \Sigma)$ , and  $\Sigma$  is given in (3.20). In the derivation of result 3.2 it was shown that  $\lambda \lambda_n \rightarrow 0$  implies:

Where  $\lambda \lambda_n \rightarrow 0$ , and  $\lambda \lambda_n \rightarrow 0$ ,  $\lambda \lambda_n \rightarrow 0$ . Since,  $\lambda \lambda_n \rightarrow 0$ , and  $\lambda \lambda_n \rightarrow 0$ , (3.18) reduces to:

Also:

Because  $\lambda \lambda_n \rightarrow 0$ , and  $\lambda \lambda_n \rightarrow 0$ . Note that the distribution of  $\hat{\beta}_n$  is known from expression (3.20) and  $\lambda \lambda_n \rightarrow 0$  where  $\lambda \lambda_n \rightarrow 0$ . From (3.31) and (3.32):

$$\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow N(0, \Sigma)$$

## 4. Estimation

### Result 4.1

Where  $\lambda$  and  $\lambda_n$  are defined in Lemma A4 in the appendix.

### Result 4.2

Updating the smoothening parameters via the equation  $\lambda_n = \lambda_n(\lambda)$  reduces  $\lambda_n$  approximately in each step until (4.1) is satisfied.

With equality if and only if (4.1) is satisfied.  
For the proofs of result 4.1 and 4.2 see section 4 in the appendix.

### Monte Carlo simulations

The updating rule in (4.1) is applied in some Monte Carlo simulations in this section. Obviously, the criterion function

defined in (3.2) and not  $\hat{f}_n$  is used. I report some simulations designed to examine the finite sample performance of the estimator defined in (2.7)-(2.9), when the value of  $\lambda$  is decided by minimisation of  $MSE(\hat{f}_n)$ .

I will consider a simple case which has one non-uniform variable and one uniform variable, both of which are binary:

with

I will consider cases with varying degrees of

The number of simulations is  $M=1000$  and the sample sizes are  $n=50, 100$  and  $200$ . I compute the average mean squared error (MSE) by  $\frac{1}{M} \sum_{j=1}^M MSE_j$ , where  $MSE_j$  is the estimated MSE in the  $j$ th simulation.

Table 4.1. MSE summary of  $\hat{f}_n$  as a function of  $n$  and  $\lambda$ .

$n$		$\lambda=0.01$	$\lambda=0.05$	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.5$	$\lambda=1$
50	0.01382	0.7706006	0.4560976	0.3011383	0.3023772	0.9545200	0.610654
100	0.00693	0.8201192	0.4742411	0.3184412	0.3181261	0.9613788	0.567116
200	0.00367	0.8336982	0.4592883	0.2851164	0.2850364	0.8953773	0.494068

$n$		$\lambda=0.01$	$\lambda=0.05$	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.5$	$\lambda=1$
50	0.01382	0.4567577	0.4553528	1.3440427	0.9854756	0.7590816	0.7549082
100	0.00693	0.4184341	0.4166647	1.7785507	1.2365734	0.9225750	0.9167837
200	0.00367	0.3198203	0.3194588	1.8936664	1.1459096	0.7906996	0.7899998

The first column,  $MSE(\hat{f}_n)$ , reports the mean square error of the frequency estimator and the other columns reports the means square error of  $\hat{f}_n$  relative to the frequency estimator.

It is clear from table 4.1 that the estimation accuracy for this particular probability distribution increases when  $n$  increases or/and when  $\lambda$  decreases. For instance, the mean square error of the estimator for  $n=50$  when using  $\lambda=0.01$  is approximately half of that using  $\lambda=1$  which correspond to least squares cross validation, i.e. the estimator in Ouyang *et al.*(2006). This was also expected from the theoretical analysis which revealed that the rate of the smoothing parameter to its optimal value is faster when either increasing  $n$  if  $f$  is uniformly distributed or decreasing  $\lambda$  if  $f$  is non-uniformly distributed.

Table 4.2 reports the average values of smoothing parameters over the  $M=1000$  simulations.

Table 4.2 Average values of smoothing parameters as a function of  $n$  and  $\lambda$ .

$n$		$\lambda=0.01$	$\lambda=0.05$	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.5$	$\lambda=1$
50		2.9e-03, .190	5.3e-04, .856	1.0e-03, .993	1.8e-03, .999	2.5e-01, .159	1.7e-01, .836
100		3.1e-04, .135	1.0e-03, .863	1.5e-04, .996	7.7e-05, .999	1.2e-01, .129	7.1e-02, .853
200		5.8e-05, .139	1.1e-05, .843	2.0e-05, .997	8.3e-06, .999	5.6e-02, .129	2.9e-02, .840

$n$		$\lambda=0.01$	$\lambda=0.05$	$\lambda=0.1$	$\lambda=0.2$	$\lambda=0.5$	$\lambda=1$
50		1.6e-01, .991	1.6e-01, .998	5.3e-01, .130	4.5e-01, .643	4.1e-01, .988	4.1e-01, .997
100		7.4e-02, .995	7.3e-02, .999	4.3e-01, .106	3.4e-01, .639	3.0e-01, .994	3.0e-01, .999
200		2.9e-02, .997	2.9e-02, .999	3.1e-01, .102	2.3e-01, .681	1.9e-01, .996	1.9e-01, .999

Table 4.2 reveals that the smoothing parameters converge to their optimal values, i.e.  $\lambda$  becomes smaller as  $n$  increases and  $\lambda$  increases when  $n$  increases for  $\lambda=0.01$ . Also,  $\lambda$  grows as  $n$  grows for both variables, i.e. the rate of a smoothing parameter towards its optimal value becomes faster when  $n$  increases if  $f$  is uniformly distributed and slower if  $f$  is non-uniformly distributed.

## 5. Empirical data: mortality and number of heart transplants

The updating rule in (4.1) is applied in praxis in this section. The data concerns the number of patients who died within 30 days of receiving a heart transplant. See “Christiansen and Morris, Bayesian Biostatistics, D. Berry and D. Stangl, editors, 1996, Marcel Dekker, Inc.”.

The data contain 3,646 heart transplants distributed over 27 months and 131 U.S. heart transplant hospitals. Two variables are used, D, the number of dead within 30 days treatment and T, the total number of heart transplants performed within this timeframe at the hospital where the patient died.

**Table 5.1** Descriptive statistics for empirical data. D= “number of dead.”

T= “number of transplants.”

	D	T
Min	0	1
Median	2	18
Max	18	152
Number of unique realisations	12	55

There are 12 unique values for D and 55 for T in the data. Descriptive statistics for the data can be found in table 5.1.

Table 5.2 contains the estimated smoothing parameters obtained by use of (4.1) and by using different values of  $\alpha$ .

In the software package np, which is part of the statistical programming environment R, the function “npudens” can be found, which can be used to calculate smoothing parameters using different criterion functions and different estimators. By choosing the kernel function “Aitchison and Aitken,” which is the same as the one used in the estimator (2.8) except for the transformation

—, and by choosing, “Bandwidth Selection Method” to be “least square cross validation,” which is the same as choosing the criterion function  $\phi$  the following is obtained;  
 . By use of the transformation — it can be seen from the column with values  $\alpha$  of table 5.2 that this estimate is identical to the one obtained by use of the updating rule (4.1).

**Table 5.2** Smoothing parameters as a function of  $\alpha$

$\alpha$	$\hat{\alpha}_D$	$\hat{\alpha}_T$	$\hat{\alpha}_{D+T}$	$\hat{\alpha}_{D/T}$	$\hat{\alpha}_{T/D}$	$\hat{\alpha}_{D \cdot T}$
0, .335	0, .780	0, .997	0, .999	.651, .062	.059, .768	

$\alpha$	$\hat{\alpha}_D$	$\hat{\alpha}_T$	$\hat{\alpha}_{D+T}$	$\hat{\alpha}_{D/T}$	$\hat{\alpha}_{T/D}$	$\hat{\alpha}_{D \cdot T}$
0, .997	.021, .999	.933, 0	.909, .134	.097, .996	.168, .999	

## 6. Conclusion

This paper concerns the estimation of multi-dimensional discrete random variables. The probability distribution for a multi-dimensional discrete random variable can be estimated by the maximum likelihood estimator. When the number of observations is small relative to the number of parameters to be estimated, it may be advantageous to use a method which smoothes the parameter estimates through a weight function. This is supported by the fact that the maximum likelihood estimator is not efficient when at least one of the variables is uniformly distributed.

The paper examines an estimator which smoothes the parameter estimates (least square cross validation), including the deduction of the asymptotic probability distribution for a smoothing parameter which corresponds to the only uniformly distributed variable.

The paper also shows why smoothing parameters, which correspond to uniformly distributed variables, have a positive probability of not converging to their optimal value when using least square cross-validation. Finally, a



criterion function is suggested so that smoothening parameters asymptotically converge to their optimal values.

## References

- Aitchison J. og Aitken C.G.G. (1976) Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413-420.
- Farebrother, R.W. (1984). AS 204 The Distribution of a Positive Linear Combination of Random Variables, *Journal of the Royal Statistical Society (Series C)*, Vol. 33, No. 3, pp. 332-339.
- Ouyang D, Li Q. og Racine J.(2006) Cross-validation and the estimation of probability distributions with categorical data. *Nonparametric Statistics*. Vol. 18. No. 1, January 2006, 69-100.
- Hall P., Racine J. and Li Q.(2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468), 1015-1026.
- Ruben, H. (1962). Probability content of regions under spherical normal distributions. IV The distribution of homogeneous and non-homogeneous quadratic functions of normal variables, *Annals of Mathematical Statistics*, Vol. 33, pp. 542-570.
- Sheil, J., & O'Muircheartaigh, I. (1977). AS 106 The Distribution of Non-Negative Quadratic Forms in Normal Variables, *Applied Statistics*, Vol. 26, Number 1.

## Appendix

### Lemma A1.1

Where  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{Z}$

### Proof:

By definition  $\mathbf{X}$ , so the proof is finished when I have proven (i) and (ii)

Proof of (i):

Proof of (ii):

Where

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right) \\
& = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right) \\
& \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right)
\end{aligned}$$

Thus, the identically distributed assumption implies

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right) \\
& \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right) \\
& \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right)
\end{aligned}$$

Q.E.D.

## Lemma A 1.2

Where  $\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right)$  and  $\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right)$ .

**Proof:**

Note that

, where  $\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h_n}} \left( \frac{1}{\sqrt{h_n}} \sum_{j=1}^n \frac{1}{\sqrt{h_n}} \right)$ , hence (A1.5) follow from (A1.4) and (A1.6), so only need to prove (A1.6).

The proof is finished when (A1.7) and (A1.8) are proven:

Proof of (A1.7):

Where

Now I prove (A1.8):



Where I used (A1.1) and (A1.2). (A1.8) follows from (A1.9) and (A1.3).  
Q.E.D.

### Lemma A1.3

Decompose into two sets, , , where and are the sets indicating respectively the uniformly<sup>16</sup> and non-uniformly distributed variables.

is unrelated to , that is,

Where  
and is the set of subsets of of cardinality at least 1, i.e.  
, where is the power set of , i.e. . Note the convention,  
and ;

### Proof

Using lemma 1.3.1

where . and satisfy the property that any set they contain always have a number belonging to that set corresponding to a uniformly distributed variable, i.e., . From lemma A1.3.2 then  
if or . Because if or (A1.11) reduces to (A1.10).  
Q.E.D.

### Lemma A1.3.1

Where is the set of subsets of of cardinality at least 1 and and are defined in lemma 1.3.

### Proof

From lemma A1.1

<sup>16</sup> The variables belonging to  $U$  is defined in (2.1) page 3.

The lemma follows from this.

**Lemma A1.3.2**

If  $\mathbf{X}$  then  $\mathbf{Y}$ , where

**Proof**

By assumption  $\mathbf{X}$  and from this I calculate

Where the second last equality used that  $\mathbf{X}$  is independent of the other random variables because  $\mathbf{X}$  is uniformly distributed and the last equality used that  $\mathbf{Y}$  again because  $\mathbf{Y}$  is uniformly distributed.

Q.E.D.

**Lemma A1.4**

Where

and  $\mathbf{X}$  ;  $\mathbf{Y}$  ,  $\mathbf{Z}$  .

If  $\mathbf{X}$  then  $\mathbf{Y}$

**Proof**

$\mathbf{X}$  is an unbiased estimator of  $\theta$ , i.e.  $E[\mathbf{X}] = \theta$ , this proves  $\mathbf{X}$  is unbiased. Lemma A1.3.2 and the assumption  $\mathbf{X}$  implies  $\mathbf{Y}$ . The variance of  $\mathbf{X}$  is therefore

$\text{Var}(\mathbf{X}) = \sigma^2$ , which proves (A1.17). Note that (A1.17) is implied by (A1.15) and (A1.16). The only thing needed to derive the rate of  $\mathbf{X}$  is therefore the rate of its variance, i.e. to show that  $\sigma^2 > 0$ ,

, thus if .

Where the last equality comes from

, and

. To see that the first equality equals zero write

Where the second last inequality follows from  
independent of

being  
and the last from  
, since

is uniformly distributed. That

is derived in a similar manner. Then from (A1.19)

where .

This is true because ,

and since

for any value of  $\lambda$ ,

Finally, from (A1.18) and (A1.20),

This means that the mean square error of  $\hat{\theta}$  is of the order  $O(n^{-1})$  which implies that

Q.E.D.

### Theorem A2

Let  $\lambda_1, \lambda_2$  be the smoothening parameters corresponding to the respectively non-uniformly and uniformly distributed variables. The global (unique)<sup>17</sup> minimum point is

With minimum

### Proof

From lemma A1.3

and if  $\lambda_1 = \lambda_2$  then

, i.e. the estimator turns in to the frequency estimator

, for  $\lambda_1 = \lambda_2$ , and the frequency estimator is unbiased. This proves (A2.1) and

(A2.2).

The only thing left to prove is that  $\hat{\theta}$  is the unique minimum point, i.e. to prove (A2.3).

Lemma A2.1 show that  $\hat{\theta}$  is uniformly distributed with respect to  $\lambda$  is a necessary condition for (A2.3) to be true. Thus, if there is only one variable, i.e.  $d=1$ , the proof is finished. So, assume that there is more than one variable. There are two cases either  $\lambda_1$  is the only marginal uniform distributed variable or there are more than one marginal uniform variable.

Assume that  $\lambda_1$  is the only marginal uniform distributed variable, then lemma A2.2 shows that  $\hat{\theta}$  is uniform with respect to  $\lambda_1$  is a necessary condition for (A2.3) to be true. This completes the proof for this case.

<sup>17</sup> The minimum point is unique with respect to  $\lambda$  which have to be  $\lambda_1 = \lambda_2$ , but  $\lambda_1$  can be any value.

Assume that more than one of the variables are uniformly distributed with respect to a subset of  $\mathcal{V}$ , i.e. assume that  $\mathcal{V}_1$  is uniform with respect to  $\mathcal{V}_1$ ,  $\mathcal{V}_2$  is uniform with respect to  $\mathcal{V}_2$ . Note that  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$  (see (A2.7)). There are two cases either  $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$  or  $\mathcal{V}_1 \cup \mathcal{V}_2 \subsetneq \mathcal{V}$ . Assume  $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$ , since  $\mathcal{V}_1$  is uniform with respect to  $\mathcal{V}_1$ ,  $\mathcal{V}_2$  satisfy (A2.3). Repeatedly applying Lemma A2.4 shows that

$\mathcal{V}_1$  is a necessary condition for (A2.3) to be true. (A2.4) shows that  $\mathcal{V}_2$  is uniform with respect to  $\mathcal{V}_2$ ,  $\mathcal{V}_1$  is a necessary condition for (A2.3) to be true, this finishes the proof for this case. Now, assume that  $\mathcal{V}_1 \cup \mathcal{V}_2 \subsetneq \mathcal{V}$ , i.e.  $\mathcal{V}_1 \cup \mathcal{V}_2 \neq \mathcal{V}$ ; the proof is by induction. Note that  $\mathcal{V}_1$  is uniform with respect to  $\mathcal{V}_1$  is necessary for (A2.3) to be true by Lemma A2.2, this is the initial step of the induction proof. The induction step is Lemma A2.3 and the proof now follows from the induction principle, i.e.  $\mathcal{V}_1$  is uniform with respect to  $\mathcal{V}_1$ ,  $\mathcal{V}_2$  is necessary for (A2.3) to be true. Q.E.D.

### Lemma A2.1

$\mathcal{V}_1$  is necessary for (A2.3) to be true.

#### Proof:

A minimum point satisfy

Summing over all variables but  $\mathcal{V}_1$  gives:

Q.E.D.

### Lemma A2.2

Define the sets  $\mathcal{V}_1, \mathcal{V}_2$  by

$\mathcal{V}_1$ , i.e.  $\mathcal{V}_1$  is the set that indicates the variables that are not marginally uniformly distributed and  $\mathcal{V}_2$  is the set that indicates the rest of the variables  $\mathcal{V} \setminus \mathcal{V}_1$ .

Assume that  $\mathcal{V}_1$  is uniformly distributed with respect to  $\mathcal{V}_1$ , and  $\mathcal{V}_2$ , then

is necessary for (A2.3) to be true.

#### Proof:

From Lemma A2.1  $\mathcal{V}_1$  is necessary for (A2.3) to be true and inserting this in (A2.5) and summing (A2.5) over  $\mathcal{V}_2$  gives

Q.E.D.

### Lemma A2.3

Assume



is necessary for (A2.3) to be true, then

, is necessary for (A2.3) to be true.

**Proof:**

From Lemma A2.1 is necessary for (A2.3) to be true and inserting this in (A2.5) and summing (A2.5) over gives

By use of expression (A2.10) it can be shown (see Lemma A2.3.2) that

Inserting the right hand side of (A2.13) into (A2.12) gives

Note that

With equality if and only if . Thus, (A2.11) is necessary for (A2.3) to be true.  
Q.E.D.

### **Lemma A2.3.2**

Assume (A2.10) is true then (A2.13) is true.

**Proof**

Now use expression (A2.10)

Q.E.D.

**Lemma A2.4**

Assume  $\alpha \in \mathcal{A}$ ,  $\beta \in \mathcal{B}$ , then

$\alpha, \beta$  is necessary for (2.3) to be true.  
The proof is very similar to Lemma A2.3.

**Lemma A3**

Where

**Proof**

From (A3.5) and (A3.11) it is clear that

Where  $\alpha \in \mathcal{A}$ ,  $\beta \in \mathcal{B}$  is defined in Lemma A3.2. (A3) follows from this.

Q.E.D.

**Lemma A3.1**

**Proof**

(A3.5) follows from (A3.1)-(A3.4).

Proof of (A3.1):

$\hat{\mu}_h$  is defined in the same way as  $\hat{\mu}_h$  (see (A1.12) not (A1.10)) except that  $\hat{\mu}_h$  is substituted with an unbiased estimator hereof  $\hat{\mu}_h$ . The expression then follows from the decomposition in Lemma A1.3. Note that  $\hat{\mu}_h$  is defined in Lemma A1.3 and  $\hat{\mu}_h$  in Lemma A1.4.

Proof of (A3.2):

$\hat{\mu}_h$ , where  $\epsilon_h$  (1) is a stochastic variable which follows from Lemma A1.4.

Proof of (A3.3):

The last equality used Lemma A1.3.  $\hat{\mu}_h$  is a sum of functions of  $\hat{\mu}_h$  that is of smaller order when minimising  $\hat{\mu}_h$ . First note that Lemma A1.4 gives  $\hat{\mu}_h$ . Hence,

Theorem A2 showed that  $\hat{\mu}_h$ , and that  $\hat{\mu}_h$  is a necessary condition for minimisation of the integrated squared bias, thus clearly any smoothing parameter that does not correspond to a variable that is uniformly distributed with respect to all variables, i.e. does not satisfy (2.1), and minimises  $\hat{\mu}_h$  converges to zero in probability, i.e.  $\hat{\mu}_h \rightarrow 0$ .

Proof of (A3.4):

Where  $\hat{\mu}_h$ ,  $\hat{\mu}_h$ , and  $\epsilon_h$  (1) is a stochastic variable which follows from Lemma A1.4. Expression (A3.4) follows from this.

Q.E.D.

### Lemma A3.2

Where

,

and

#### Proof

(A3.11) follows from (A3.7)-(A3.10).

(A3.7) is derived similar to the decomposition in lemma 1.3. Note that two terms are missing on the right hand side of (A3.7) , i.e. the terms . They do not depend on and is deliberately left out of the expression.

Proof of (A3.8)

Expression (A3.8) follows from this.

Proof of (A3.9):

where , expression (A3.9) follows from this.

Proof of (A3.10)

Expression (A3.10) follows from this.  
Q.E.D.

**Lemma A3.3**

---

**Proof**

, and  $\frac{\partial \mathcal{L}}{\partial \lambda}$  with minimum  $\lambda^*$ , clearly  $\lambda^* \geq 0$  and  $\lambda^* \leq 0$  with maximum  $\lambda^*$ .  
Q.E.D.

**Lemma A3.4**

, where  $\lambda^* = 0$ ,

**Proof**

From lemma A3,  $\lambda^* = 0$  is obvious, thus, need to prove that  $\lambda^* < 0$ . From  $\lambda^* = 0$ , it is clear that  $\lambda^* = 0$ , hence,  $\lambda^* = 0$  finishes the proof.

Note that  $\lambda^* = 0$  follows from lemma 2.3.  
Thus,

Hence,  
 Q.E.D.

**Lemma A3.5**

**Proof**

Where  $\dots$  and  $\dots$ . Expression  $\dots$  follows from this.

**Lemma A3.6**

Let  $\dots$  be an identical and independent distributed sample, then

$\dots$ . Where  $\dots$ ,  $\dots$  and  $\dots$ .

**Proof**

Observe that  $\dots$ , hence, (A3.16) follows from the central limit theorem.  
 Q.E.D.

**Lemma A3.7**

**Proof**

Using Lemma A1.3:

Where  $\mathbf{A}$  is a  $n \times n$  cross matrix with entries  $a_{ij}$  if  $i=j$  and  $0$  if  $i \neq j$ ,  
 $\odot$  and  $\odot$  is the elementwise multiplication operator. Similar to theorem 2,  
 although easier one can show that  $\mathbf{A} \odot \mathbf{B} = \mathbf{C}$  with equality if and only  
 if  $\mathbf{A} = \mathbf{C}$ .  
 Q.E.D.

**Lemma A3.8**

**Proof**

Hence,

Q.E.D.

**Result 4.1**

Define  $\mathbf{A}$ . Then:

Where  $\mathbf{A}$  ;  $\mathbf{B}$  ;

$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$ ;  $\frac{\partial \hat{\mu}_i}{\partial \mu_j} = 0$ ; and  $\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$  and  $\frac{\partial \hat{\mu}_i}{\partial \mu_j} = 0$ .

**Proof**

From Lemma A1.3.1 and Lemma A1.2 and the definition of  $\hat{\mu}_i$  in (3.1):

Note that  $\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$ , So:

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

$$\frac{\partial \hat{\mu}_i}{\partial \mu_j} = 0$$

Which mean that:

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

$$\frac{\partial \hat{\mu}_i}{\partial \mu_j} = 0$$

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

Where the last equality used the definition of  $\hat{\mu}_i$  and  $\mu_i$  in result 4.1. Expression (4.1) can be obtained by setting expression ( ) equal to zero and rearrange. Q.E.D.

**Result 4.2**

Updating the smoothening parameters via (4.1) reduces  $\hat{\mu}_i$  approximately in each step until (4.1) is satisfied.

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

With equality if and only if (4.1) is satisfied.

**Proof**

An approximation to the change in  $\hat{\mu}_i$  from updating the smoothening parameters via (4.1) can be obtained by a linear Taylor approximation:

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

Updating via (4.1) gives:

$$\frac{\partial \hat{\mu}_i}{\partial \mu_i} = 1$$

$$\frac{\partial \hat{\mu}_i}{\partial \mu_j} = 0$$

With equality if and only if (4.1) is satisfied. This follows from Q.E.D.



### Lemma A4.3

and unless if is uniformly distributed with respect to .

#### Proof

is defined in result 4.1: . Inequality (A4.8) follows directly from (A4.9), thus I will prove (A4.9).

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}}$$

Similarly:

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}}$$

So:

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}}$$

Because, and and with equality iff. is uniformly distributed with respect to . Since, clearly also . To complete the proof observe that:

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}}$$

Because, — .  
Q.E.D.

### Lemma A5

Let be defined as:

Note that is a function of , . Especially . Then:

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{1 - \frac{1}{n}}}$$

Where — — and — — and is a stochastic variable which has a chi-square distribution with degrees of freedom.

Assume that is uniformly distributed with respect to , then:

Assume that  $\mathbf{X}$  is uniformly distributed with respect to  $\mathbf{P}$ , then:

**Proof**

I will first prove

$$\begin{aligned} & \int \mathbf{X} d\mathbf{P} = \int \mathbf{X} d\mathbf{P} \\ & \quad - \end{aligned}$$

Where  $\mathbf{X} =$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

$$-$$

Note that the last equality used  $\mathbf{X}$  in lemma A5.2.2 and the fact that  $\mathbf{X}$  is uniformly distributed with respect to  $\mathbf{P}$ . Q.E.D.

For the proof of  $\mathbf{X}$  I define:

Where  $\mathbf{X} =$  and  $\mathbf{X} =$

Similar to the derivations that led to (5.6), clearly:

Where  $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$  and  $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$ .

Where the last equality used  $\mathbf{A} = \mathbf{B}$  in lemma A5.2.2 and the fact that  $\mathbf{A} = \mathbf{B}$ .  
Q.E.D.

For the proof of (A5.4) note that the assumption:  
 $\mathbf{X}$  is uniformly distributed with respect to  $\mathbf{A}$   
 Thus,  $\mathbf{X} \sim \mathbf{A}$ , hence

And  $\mathbf{X} \sim \mathbf{A}$  follows from the multivariate Central Limit Theorem.  
Q.E.D.

For the proof of (A5.5) I will first show that:

Where  $n$  is the number of support points of  $\mathbf{X}$ ;  $\mathbf{I}_n$  is the identity matrix or unit matrix of size  $n$  and  $\mathbf{D}$  is a positive definite symmetric block diagonal matrix having main diagonal blocks square matrices,  $\mathbf{D}_{ii}$ , and off-diagonal blocks are zero matrices.  $\mathbf{D}$  is of size  $n \times n$ .

## Proof

The third equality in (A5.9) is trivial and the second follows from the definition of a bijective function:

Then:

$$\begin{aligned} \mathbf{A} &= \mathbf{B} \\ \mathbf{A} &= \mathbf{B} \end{aligned}$$

To prove the first equality in (A5.9) I calculate:



The variance matrix of  $\mathbf{y}$  is:

So:

Since the variance matrix is a diagonal matrix expression (A5.20) show that the elements of  $\mathbf{y}$  are asymptotically uncorrelated and because  $\mathbf{y}$  also have a multivariate normal distribution each element in  $\mathbf{y}$  are independent of each other and normally distributed. For instance,  $y_i$  is normally distributed with variance  $\sigma^2$ , so:

Where  $z_i$  are independent standard normal variables. Finally from (A5.19):

The vector  $\mathbf{y}$  has size  $n$ , so:

Where  $z_i$  is a stochastic variable that is chi-square distributed with 1 degree of freedom. Each  $z_i$  in the sum is independent from the others which follow from  $z_i$  being independent standard normal variables.  
Q.E.D.

To complete the proof of (A5.5) I will finally show that:

Expression (A5.24) is implied by:

Where  $f^{-1}$  is the inverse function of  $f$  defined in (A5.25) is true if and only if exactly  $m$  of the eigenvalues equal  $\lambda$ .

The eigenvalues of the matrix  $\mathbf{A}$ , can be found as the solution of the equation:

Where  $P(\lambda)$  is the characteristic polynomial. I will show that:

Expression (A5.25) follows from (A5.27) by noting that  $\mathbf{A}$  is symmetrical, hence, the algebraic multiplicity, which is  $m$  for each eigenvalue, equal the geometric multiplicity of each eigenvalue.

#### **Proof of (A5.27)**

The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are block diagonal matrices and  $\mathbf{C}$  is therefore also a block diagonal matrix:

with main diagonal blocks square matrices:

. Because of this block diagonal structure, the characteristic polynomial can be written:

The proof is therefore finished if:

Note that (A5.31) in Lemma A5.3 show that  $\mathbf{A}$ , consequently:

So,

Q.E.D.

**Alternative Proof of (A5.25)**

Using the definition of  $\mathcal{L}$  and  $\mathcal{M}$  in respectively (A5.30) and (A5.13) gives:

Where the last equality used (A5.31) in Lemma 5.3. This show that:

Thus, from expression (A5.17):  $\mathcal{L}(\mathcal{M}(x)) = \mathcal{L}(x)$  it is clear that  $\mathcal{L}(\mathcal{M}(x)) = \mathcal{L}(x)$  and:

Expression (A5.18) then show that  
Q.E.D.

**Lemma A5.2**

**Proof**

Which prove the first equality in (A5.7). For the second equality in (A5.7), it is easy to see that  $\mathcal{L}(\mathcal{M}(x)) = \mathcal{L}(x)$ , with

To prove (A5.8) I calculate:

Where

Thus,

Which prove (A5.8).  
Q.E.D.

**Lemma A5.3**

Let

Where  $I_n$  is the identity matrix of size  $n$  and  $\mathbf{A}$ .

Then:

Where  $\mathbf{A}^T$  and  $\mathbf{A}$  and

**Proof**

$\mathbf{A}$  is the matrix for which

and  $\mathbf{A}^T$  defined in (A5.30) satisfy equation (A5.32). To see this note that if  $\mathbf{A}$  is defined as in (A5.30) then all diagonal elements of  $\mathbf{A}^T \mathbf{A}$  equals:

and all non-diagonal element in  $\mathbf{A}^T \mathbf{A}$  equals:

So, if  $\mathbf{A}$  then it must be the case that:  
and  $\mathbf{A}^T$ , which in both cases are equivalent to:

Solving this equation yields:  $\mathbf{A}^T = \mathbf{A}$ . This proves (A5.30). To prove (A5.31), notice that,

when  $\mathbf{A}$  is defined as in (A5.30).

Q.E.D.

# Paper 4: Consumption effects of a tax on saturated fat in foods

---

Aslak Hedemann Hansen<sup>18</sup>

University of Copenhagen, Institute of Food and Resource Economics

---

<sup>18</sup> The author would like to thank the financial supporter; The Danish council for strategic research.



# Consumption effects of a tax on saturated fat in foods

---

## **Abstract.**

The purpose of this paper is to examine if a tax on saturated fat can improve the Danes' diet, while taking the possible population heterogeneity of the demand response to the tax into account.

Great heterogeneity regarding price responsiveness within the Danish population underlines the importance of the household specific model approach, and indicates that consumers should be treated accordingly when considering the development of new food taxes. The analysis reveals that a likely outcome of the tax will be an improved diet for the majority of the population.

**Keywords:** Household Specific Prices, Price Elasticity, Healthy Diet, Public Policy.

## 1. Introduction

The prevalence of being overweight and obese is increasing at a rapid pace. According to WHO (2006) figures, approximately 1.6 billion adults were overweight, and at least 400 million adults were obese globally in 2005, and by 2015, approximately 2.3 billion adults will be overweight and more than 700 million will be obese. The EU conforms to this trend, according to the European Commission (Platform Charter, 2005); European Union citizens are experiencing, “a sustained, acute EU-wide increase in overweight and obesity”. The increase is particularly severe for children for whom the estimated incidence of overweight was 30% in 2006 (European Commission, 2007). Improving the diet quality amongst the European population, as well as other measures designed to contain or reverse current obesity trends, has consequently become a prioritised matter and is now on the political agenda in the EU.

Evidence links obesity with a diet containing a relatively high intake of energy-dense foods and saturated fatty acids (SFA). Therefore, policies that aim at limiting SFA, or decreasing the SFA to unsaturated fatty acids ratio, are considered likely to improve obesity prevalence rates<sup>19</sup> (WHO, 2003a). This evidence was translated into recommendations and guidelines in the WHO Global Strategy on Diet, Physical Activity and Health (WHO, 2003b, 2004), which was adopted in May 2004. The WHO recommends, “moving from saturated animal-based fats to unsaturated vegetable-oil based fats” (WHO, 2003b), to “limit energy intake from total fats and to shift fat consumption away from saturated fats to unsaturated fats” (WHO, 2004). Recently, the European Food Safety Authority (EFSA, 2009; p.2) put forward a recommendation concerning the intake of saturated fatty acids in the EU population: “The Panel recommends that SFA intake should be as low as possible within the context of a nutritionally adequate diet”. Debates about taxing unhealthy foods continue in the US and the EU and in 2009, the health commissioner for the City of New York suggested an increase in taxes notably on sugared beverages (Brownell *et al.*, 2009), whilst in Denmark, an elected tax increase of 25% on ice cream, chocolate and candy, as well as a tax on soft drinks differentiated with respect to sugar, came into effect in January 2010 (The Danish Ministry of Taxation, 2009 a).

As a means to reduce SFA intake, a new tax based on the content of saturated fat in foods has been proposed in Denmark. This fat tax is special since it not only taxes certain products, but also introduces a new element; the taxation is based on a specific nutrient contained in food products, saturated fat. The proposal was inspired by the work of a government-appointed commission (Forebyggelseskommissionen, 2009).

The base for the fat tax might be dairy products excluding milk, margarine, oils and animal fats (The Danish Ministry of Taxation, 2011a, 2011b, 2009 b).

In this paper, the potential demand effects of a saturated fat tax are analysed, taking the demand for butter and margarine as a case study. For comparison, an alternative (VAT) simulation was conducted to calculate the demand effects of increasing the prices of all the considered fat products by the same percentage. Butter and margarine contribute a significant part of the intake of SFA<sup>20</sup> as they contain a large amount of saturated fat. Consequently, the tax raises the price of butter and margarine considerably in a relative sense, and both products have close substitutes with significantly lower amounts of SFA.

---

<sup>19</sup> Recent meta-analysis based studies also gives evidence towards the hypothesis that it is more important what the saturated fatty acids is substituted with than merely reducing the SFA intake (Mozaffarian *et al.*, 2010), but emphasizes that the risk of coronary heart disease can be reduced by substitution of SFA with poly unsaturated fat.

<sup>20</sup> In Denmark, the intake of saturated fat is approximately 50% higher than recommended and 18% of SFA comes from butter and mixed butter, 14% from margarine and oils, whilst milk and cheese contribute 32% and meat 18% (Pedersen, 2010).

The demand behaviour, with regard to saturated fats (including the sensitivity of consumption with respect to price changes, e.g. induced by taxes), has been the subject of a number of other studies in the literature. For example, Yen & Chern (1992) and Koc *et al.* (2001) estimate price and expenditure elasticities for a number of fat products. Huang & Lin (2000) estimate food demand and nutrient elasticities on the basis of household survey data. Nichele (2003) estimates a set of behavioural elasticities (with respect to expenditure, price and health information) for a range of food products and links the responses in the consumption of these products to the intake of various lipids, including saturated fat, using a set of nutrient coefficients for the different food categories. Also, Dhebibi *et al.* (2007) link product consumption with nutrient coefficients in a demand system. In principle, such demand models may be used for assessing the effects of SFA-taxation. This approach has been used by two Danish studies (Smed & Jensen, 2007, Jensen & Smed, 2007), who examine the effects of various tax change scenarios on the intake of different nutrients. Common to the above-mentioned studies is the fact that they estimate food demand systems assuming a uniform statistical distribution of behavioural parameters across all individuals or households covered by the studies. A potential problem with the methodology applied in the above mentioned papers is that consumers might not be homogeneous in their response to price signals. The modern literature on program evaluation has stressed the importance of relaxing functional form and distributional assumptions including heterogeneity in the effects of the treatment as such heterogeneity is important in practice (Imbens & Wooldridge, 2009). In contrast with previous analyses on this topic, in this study such strict assumptions are not imposed. Recognising the possible variation in parameters, the individual household's price response is estimated here by unlocking more information on the expected response to the tax reform. Little knowledge of the possibility of modifying consumers' consumption behaviour with respect to foods rich in SFA through pricing policies exists, especially at the household level. This study aims to add to the knowledge of policies that are useful in reducing obesity prevalence rates by examining if policies that tax foods rich in SFA can improve the diet of the corresponding population, whilst taking the potential population heterogeneity in the response to the tax into account. The analysis is applied on the proposed Danish fat tax, which is an excellent case for this purpose. The paper is organised as follows. Section 2 describes the data applied in the analysis, whilst the methodology used in the study is introduced in section 3. The results of the analysis are presented in section 4, and finally, section 5 draws some conclusions and perspectives for policy and further research.

## 2. Data

The analysis uses household panel data based on a "nation-wide representative sample", which "is selected according to geographic and demographic criteria based on Central Statistical Bureau information" (GfK Consumer Tracking, 2010). GfK produces weekly diaries, reporting household purchases regarding quantity, price, expenditure, place of purchase, manufacturer, and other relevant product features etc. on a detailed level. The data covers the period from January 1997 to December 2004. The number of households used in the analysis is 4.059.

The present study focuses on two butter categories and three categories of margarine, and the weekly diaries are aggregated into monthly observations. In the context of healthy eating, the most important product characteristic is the fat content, i.e. whether it is made purely from milk fat (traditional butter), edible oils or a mix of these. For butter, a distinction is made between traditional butter (*butter*), and butter products which consist partially of vegetable oil (*mixed butter*). The margarines are grouped together into three categories: *Hard margarine* (used primarily for frying

and baking), which has the highest saturated fat content; *spreadable margarine* (used for spreading on bread), which contains a medium amount of saturated fat; and *minarine* (low fat margarine – also used for spreading on bread), which is especially low in fat, and hence also in saturated fat.

The fat content of the five product types are displayed in table 1. For instance, it should be noted that the mixed butter category contains a larger share of saturated fat than any of the margarine types.

**Table 1 Weight share of fat and saturated fat in the five products considered**

	Butter	Mixed Butter	Hard Margarine	Spreadable Margarine	Minarine
% share fat	81,2	80,0	82	77	35
% share Saturated Fat	51,8	39,6	28	15,3	7

### 3. Methodology

We follow the traditional neoclassical approach to consumer behaviour, assuming that rational consumers exhibit utility maximising behaviour subject to a budget constraint, and that this implies responses to changes in prices and income. Furthermore, we assume that the consumption of fat products (butter and margarine) are weakly separable from the consumption of other goods, i.e. the marginal rate of substitution between any two goods within this fat composite is unaffected by the consumption of any good outside this group of fat products (Deaton & Muellbauer, 1980a).

#### 3.1 Empirical model

The basic ingredient in the empirical model is the Almost Ideal Demand System (AIDS) model of Deaton & Muellbauer (1980b). An AIDS demand system is estimated for each household. Each household  $h$  purchases a subset of the five goods,  $\{j \in J_h\}$ , where  $J_h$  is the set of goods purchased by household  $h$ . The AIDS model assumes a particular expenditure function, which gives rise to the budget share equations for household  $h$  in an  $I$ -good system:

where  $w_{ih}$  is household  $h$ 's budget share associated with the  $i$ 'th good,  $i \in I$ ,  $\alpha_i$  is the intercept term in the  $i$ 'th budget share equation,  $\beta_j$  is the slope coefficient associated with the price of the  $j$ 'th good ( $p_j$ ) in the  $i$ 'th budget share equation. To incorporate seasonality and trend in the consumption data we follow Ghysels *et al.* (2001) and augment the AIDS model with trigonometric variables and a linear trend variable,  $\tau$ . Thus,

where  $\alpha_i$  and  $\beta_i$  represent parameters on the trigonometric variables, and  $\gamma_i$  is the parameter on the time trend.  $E_i$  is the total expenditure on the system of goods given by  $E_i = \sum_{j=1}^J p_j q_{ij}$  in which  $q_{ij}$  is the quantity demanded of the  $i$ 'th good.  $P_i$  is the price index defined by

.

Adding up is imposed on the system:  $\sum_{i=1}^I E_i = E$ ,  $\sum_{i=1}^I \alpha_i = 0$ ,  $\sum_{i=1}^I \beta_i = 0$ , as well as Slutsky symmetry  $\alpha_i \beta_j = \alpha_j \beta_i$ . Linear homogeneity is automatically imposed, i.e.  $\sum_{j=1}^J p_j \alpha_{ij} = 0$  because of the symmetry and adding-up restrictions. The conditional Marshallian price elasticity for good  $i$  with respect to good  $j$  is calculated as:

$$\epsilon_{ij} = \frac{p_j}{E_i} \frac{\partial E_i}{\partial p_j}$$

in which  $\delta_{ij}$  is an indicator function, yielding 1 if  $i=j$  and 0 otherwise. The expenditure elasticity is given by  $\epsilon_i = \frac{E_i}{E} \frac{\partial E}{\partial E_i}$ . Estimation is accomplished by ITSUR and standard errors for the calculated elasticities are calculated using Taylor series expansions, i.e. using the delta method (Greene, 2003).

### 3.2 Construction of household level price data for the analysis

In food demand analysis it is recognised that an important initial step is to construct price data in an appropriate way, which involves two methodological issues: the construction of price data based on a basket of similar but different goods, and the issue of no purchase.

An influential approach to these issues is that of quality adjusted prices (Cox & Wohlgenant, 1986; Gustavsen *et al.*, 2008). This method was developed to deal with the first issue but incidentally also solves the issue of no purchase. Cox & Wohlgenant and Gustavsen *et al.* use data obtained from consumer expenditure surveys of a national statistical agency/bureau, and then estimate quality adjusted prices which are then used as covariates in the cross sectional empirical model of real interest. The need for adjusting prices stems from the aggregated nature of the unit prices. For example, the unit price of butter for household  $i$  should be based on the basket of butter products bought by household  $i$ , which may deviate from that of household  $j$ , due to differences in preferences, availability, etc. The quality adjusted price methodology is useful for imputing prices in a cross-sectional framework and works through the assumption that the household for which a price needs to be imputed needs to have the same goods in its basket of goods as the average household in the subgroup to which it belongs.

Because the empirical model is based on individual households, each individual household's preferences and the range of goods in a household's choice set, can be assumed to be fairly constant. Given these assumptions, we are able to estimate household-level prices directly from the data without any need for adjustment.

The second issue of no purchase is problematic as no price information is instantaneously available as a unit price calculated from the observed purchase. Cox & Wohlgenant (1986) and Gustavsen *et al.* (2008) use the OLS regression

model developed for the construction of quality adjusted prices to also construct price data for households with no purchase. The missing price for a household with no purchase is predicted using the regression model, and the information about which subgroup the household belongs to, often defined by region and time period only.

In the price construction of this paper, the problem of differing regions encountered in cross sectional data as in Cox & Wohlfgenant and Gustavsen *et al.*, is not present. In this study, the data is more detailed, which gives a number of advantages for selecting the relevant goods to be included in the good baskets more accurately.

The following describes the applied procedure for constructing price data and imputing missing price observations.

### 3.2.1 The relevant price constructed from individual preferences (prices)

For a given household, a missing price<sup>21</sup> needs to be imputed based on information concerning the retail chains visited in the given period, detailed prices of similar types of products in these chains, and the household's historically preferred types of products. This means that a good basket (and a corresponding price of this good basket) can be selected in accordance with the goods that were actually available in the choice set at the given time, at the given prices, and appropriately weighted in accordance with the historical choices, thus yielding household-specific price data for each of the five product categories<sup>22</sup>.

### 3.2.2 Prediction of missing prices

As mentioned previously, missing prices in 'no-buy' periods are a problem. This sub-section describes how the missing prices were constructed consistent with the individual household's preferences, using butter (*butter + mixed butter*) as an example<sup>23</sup>.

The preferences are modelled using a discrete choice model, namely the conditional logit model, (McFadden, 1974).

In the discrete choice model there is a total of:

alternatives to choose from. There are at most                      butter alternatives (see footnote 4) and it is assumed that the household is indifferent between any two alternatives in the set                      if the price is the same - and likewise for alternatives

---

<sup>21</sup> In months in which the household made a purchase of the considered good, i.e. in months without missing prices for the good, the average price paid by the household in the month is simply used as the relevant price in its AIDS model.

<sup>22</sup> This means that products are defined by generic product type (e.g. butter type) and by retail chain. For instance, for butter products there are up to 1194 different prices/alternatives because this is the level of accuracy in the data when each butter product type and retail chain is taken into account, i.e. that the same product can have different prices in different shops. In the case of margarine, there are up to 1872 different prices.

<sup>23</sup> There are alternative ways to predict the missing prices, e.g. one could use the same model framework but include interactions terms, such that price has different impacts on the choice probabilities for different goods etc. A different approach would be to use a non-parametric framework (Racine, 2002 and Hall *et al.*, 2004). It is considered that the approach adopted here is significantly better than a simplistic approach, e.g. taking the sample average as the missing price. While more complicated models, which for example include interaction terms, could improve performance for some households, it is likely that it would worsen for others. The implication is that improving performance not only requires that the estimates are household specific (which is the case in our estimation), but also requires household specific models which would complicate the process significantly. A non-parametric framework would require a sufficient number of observations for the models to be reliable because of the slower convergence, which is not available for all households. Thus, it is considered that the current approach represents a reasonable trade off between performance and model complexity.

in  $\mathcal{A}_j$ .

In the estimation, the individual choice sets are subsets of  $\mathcal{A}_j$ , and the choice sets used for estimation satisfy the following three requirements. First, a choice set for household  $h$  is only constructed in months where household  $h$  purchases a butter product. Second, only alternatives which can be bought in a retail chain that household  $h$  has visited in month  $t$  are eligible for entering the choice set  $\mathcal{C}_h^t$ , where  $\mathcal{A}_j$  is the set of alternatives available in shop  $j$  and  $\mathcal{S}_h^t$  is the set of shops visited by household  $h$  in month  $t$ . Finally, only alternatives which have been bought by any household in the sample during month  $t$ ,  $\mathcal{B}_t$ , are eligible for entering any households choice set in month  $t$  (because if not there is no price information). If all three requirements are satisfied for a given month, then a choice set for the month is constructed and the alternatives contained within this set are the intersection of the alternatives in the sets determined by the two last requirements. The set of household  $h$ 's choice sets constructed in this manner is

$$\mathcal{C}_h = \{\mathcal{C}_h^t \mid t \in \mathcal{T}_h\} \quad \text{where } \mathcal{T}_h \text{ are the } k \text{ months where household } h \text{ made butter purchases and } \mathcal{A}_j \text{ is the set of alternatives available in shop } j.$$

Choice probabilities are derived from four assumptions. First, a household's utility from choosing alternative  $i$ ,  $u_{hi}$ , is decomposed into an observable part,  $v_{hi}$ , and a random unobservable part,  $\epsilon_{hi}$ , i.e.  $u_{hi} = v_{hi} + \epsilon_{hi}$ . Second, the random unobservable part is i.i.d extreme value. Third, the household chooses the alternative with the highest utility. The logit choice probability,  $P_{hi}$  is the probability that household  $h$  chooses alternative  $i$  under the given assumptions which is given by:

$$P_{hi} = \frac{\exp(v_{hi})}{\sum_{i \in \mathcal{C}_h^t} \exp(v_{hi})}$$

The logit choice probability is a function of the observable utility  $v_{hi}$ , which is assumed to be linear in parameters,  $v_{hi} = \beta_i p_{it} + \gamma_i \text{type}_i + \alpha_i$ , where  $p_{it}$  is the price of alternative  $i$  and  $\text{type}_i$  is an indicator of the butter being of the traditional type, and  $\beta_i$  and  $\alpha_i$  are parameters.

In months with no purchase of a butter product a price is predicted/imputed by using (3.4) in the following manner:

$$\hat{p}_{it} = \frac{1}{N} \sum_{h \in \mathcal{H}} p_{it}$$

Where  $\hat{p}_{it}$  is the predicted price of Mixed butter in month  $t$ ;  $\mathcal{C}_h^t$  is the choice set in month  $t$  that only contains mixed butter products, which is constructed as mentioned above, except that requirement one is obviously not satisfied; The probability  $P_{hi}$  is calculated as in formula (3.3) and  $p_{it}$  is the price of alternative  $i$  in month  $t$ . The predicted price of traditional butter is calculated in a similar fashion.

### 3.2.3 Advantages of individual preferences (prices)

The quality adjusted price methodology tries to find the most appropriate unit price for a basket of goods at the household level, and this when the considered baskets of goods are very heterogeneous across households.

The detailed data used in this study makes the baskets of goods within households much more homogenous (relative to the baskets of goods across households considered in Cox & Wohlgemant and Gustavsen *et al.*) and eliminates the need to adjust for regions. Finally, it allows the determination of the unit price consistent with preferences, i.e. to select the unit price of the basket of goods used in the empirical model in section 3.1 according to the most likely goods in the basket. Consequently, the unit price of the basket of goods used in the empirical model is expected to be closer to the true unit price when using individual preferences (prices), as in this study, than when using more aggregated prices. This has a direct implication in terms of more accurate price elasticity estimates of the empirical model in section 3.1, upon which the primary conclusions are based.

## 4. Results

The first part of this section presents econometric estimation results, in terms of price elasticities, whereas the second part presents and discusses the results of a price simulation experiment reflecting a tax on saturated fats.

### 4.1 Distribution of own price and cross price elasticities

In the discussion of the Danes' price sensitivity, initially the focus will be on measures of central tendency for the own price elasticities<sup>24</sup>.

The average own price elasticity of mixed butter is -1.22 and the mode value is -0.95. Although households with an own price elasticity equal to the mode value have inelastic demand, the majority of the sample has an own price elasticity (in absolute value) larger than one, i.e. elastic demand. For the other goods, the majority of the population exhibits elastic demand but there is also a significant part (more than one quarter) with inelastic price elasticity for all the considered goods.

**Table 3 Measures of central tendency for own Price Elasticity estimates and model fit statistics**

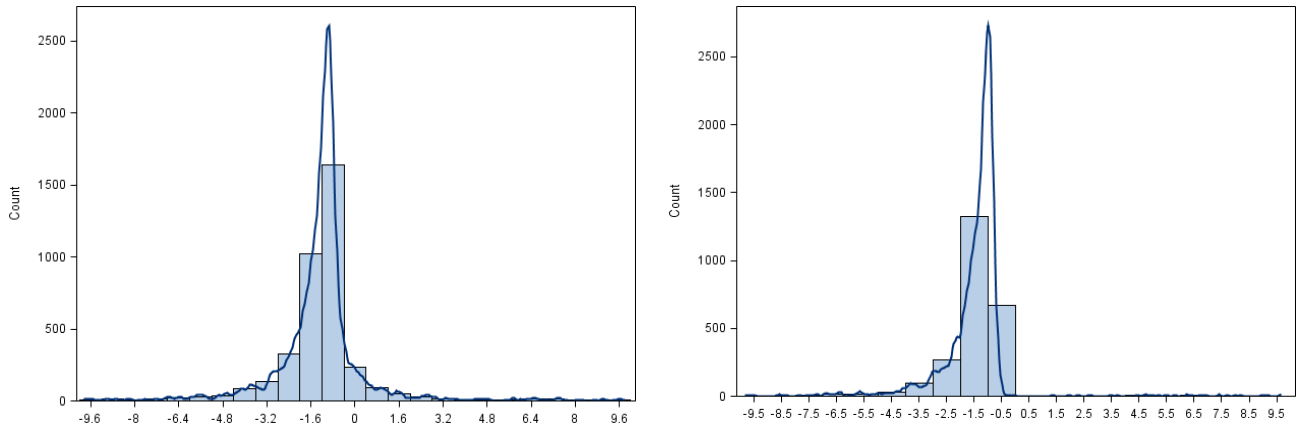
Good	Mode	Mean	Median	R <sup>2</sup>		Adj. R <sup>2</sup>	
				P25	P75	P25	P75
Butter	-1.04	-2.0	-1.79	0.24	0.63	0.06	0.51
Mixed Butter	-0.95	-1.22	-1.12	0.29	0.65	0.12	0.52
Hard Margarine	-1.13	-1.20	-1.21	0.32	0.67	0.16	0.55
Spreadable Margarine	-1.09	-1.77	-1.48	0.25	0.64	0.12	0.55
Minarine	-1.19	-1.23	-1.27				

Mode, Mean and Median statistics for samples where all estimates are included except extreme cases and percentiles for model fit statistics R<sup>2</sup> and Adjusted R<sup>2</sup>.

<sup>24</sup> An appendix containing additional statistics is available from the author upon request.



**Figure 1 Distribution of Mixed Butter's own price elasticity. Left: All estimates. Right: Significant estimates.**

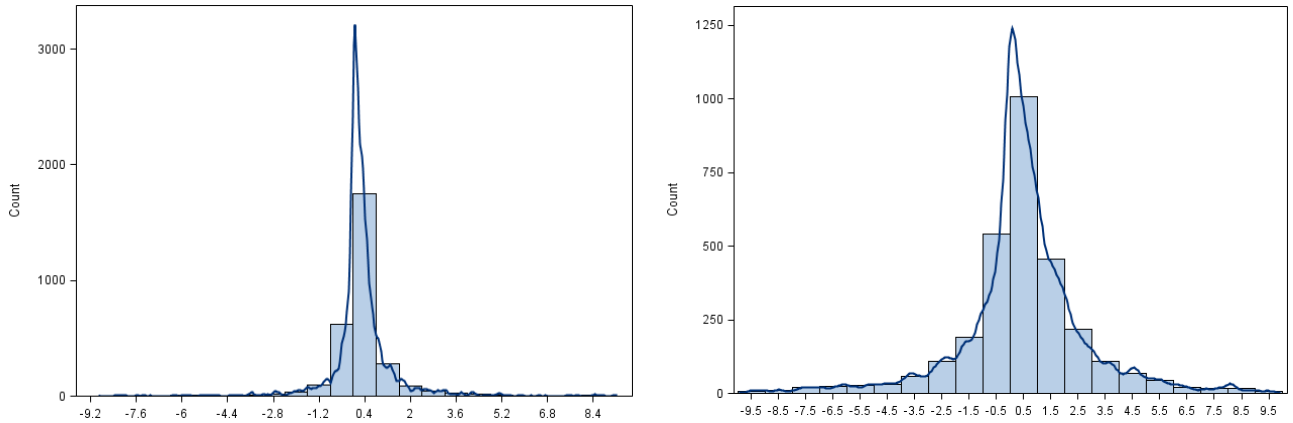


The majority of the own price elasticities are negative as theory predicts. If the insignificant price elasticities are ignored, the result is almost no positive price elasticities, which is especially clear for Mixed Butter (figure 1). The distribution of the own price elasticities for sub-samples of the estimates were also examined where the subsamples were based on different cut-offs for the minimum number of observations used in each estimation. This exercise showed that the shape of the distribution of own price elasticities did not change dramatically, hereby suggesting that the shape was insensitive to this cut-off.

The households have on average the largest own price elasticities for *butter* and *spreadable margarine* (Table 3) and *ceteris paribus* are most price sensitive with respect to these goods. It is also noteworthy that a large proportion of the households are relatively insensitive to the price when considering the healthier alternative *minarine*. However, it might raise some concern that the sample households are also relatively insensitive to price changes when considering hard margarine, a good containing a large percentage of saturated fat.

As *butter* and *mixed butter* have similar attributes, it can be expected that they will be economically close substitutes. The distribution of the cross-price elasticities between these two products is illustrated in figure 2. The demand sensitivity of *mixed butter* with respect to the price of traditional *butter* appears to be larger (mean 1.14, median 1.21, mode 0.82) than the reverse (mean 0.61, median 0.45, mode 0.36).

**Figure 2 Distribution of Cross Price Elasticity. Left: ( ) % Mixed Butter demand change per %Butter price change. Right: Vice Versa.**



Most of the mass for the distribution of  $\epsilon_{j,b}$ <sup>25</sup> is located in the interval (0;0.8) while the estimates of  $\epsilon_{b,j}$  (see footnote 7 for a definition) are much more spread out. The distributions of the cross price elasticities between *butter* and *mixed butter* indicate that a significant part of the population is more prone to substitute in favour of butter when the price of mixed butter increases than vice versa. This has an important implication for the saturated fat tax, as part of the population *ceteris paribus* has a higher probability for choosing *butter* relative to *mixed butter* after the tax change than before.

Hence, we observe much more heterogeneity when substitution is in favour of *butter* as a result of *mixed butter* price increases than vice versa. The picture was the same for the own price elasticities, i.e. the households have much more heterogeneous responses (changes in purchases of butter) when the price of butter increased. If the insignificant estimates are ignored, then most of the negative cross-price elasticity estimates disappear, hereby confirming the hypothesis of the goods being substitutes.

After Edgerton (1997) and Rickertsen (1998), the unconditional elasticities are calculated from estimated conditional elasticities:

in which  $\epsilon_{j,b}^h$  is household  $h$ 's conditional elasticity estimated from the AIDS demand system;  $\epsilon_{b,j}^h$  is household  $h$ 's expenditure elasticity estimated from the AIDS demand system;  $\alpha_j$  is the  $j$ 'th good's share of the budget for butter and margarine. Finally  $\epsilon_{b,j}^h$  is household  $h$ 's elasticity indicating the percentage change in aggregated expenditure for butter and margarine, when a one percentage change in the aggregate price index  $P_{b,j}$  of these goods is observed<sup>26</sup>.  $\epsilon_{b,j}^h$  is obtained as the estimate of the linear regression:

<sup>25</sup>  $\epsilon_{j,b}$  is % Mixed Butter demand change per % Butter price change and  $\epsilon_{b,j}$  is % Butter demand change per % Mixed Butter price change.

<sup>26</sup> The mean and standard error of the distribution of  $\epsilon_{b,j}$  equals 0.04, 0.35 and its quartiles are -0.07, 0.02 and 0.14.

Where  $\hat{\alpha}_i$  and  $\hat{\beta}_i$  are estimates from the AIDS analysis. It should be noted that the price index is derived from economic theory (Deaton & Muellbauer, 1980a), and that by using the estimated parameters  $\hat{\alpha}_i$ ,  $\hat{\beta}_i$  as weights, the price index is constructed such that it is most likely to be true conditional on the observed data and on the economic theory/model being true, i.e. it is the MLE.

## 4.2 Effect of a tax on saturated fat

Partial analysis of the own price and cross price elasticities shows that the population is quite price responsive, indicating that price instruments can be useful, but *ceteris paribus* analysis can only provide limited information because own- and cross-price effects interact. To fully understand the impacts of a saturated fat tax on the expected demand, the way in which the different price elasticities interact with the household specific tax induced price changes needs to be considered.

We will examine the effect of implementing a tax of 1 DKK pr. Kg of saturated fat. The effect of the tax on the population purchase behaviour depends on its price sensitivity measured by price elasticities and by the household specific price changes induced by the tax. Table 4 shows the impact of a 1 DKK/kg saturated fat tax on the consumer prices of the five goods.

**Table 4 Distribution of sample Households (relative) percentage price change induced by Tax of 1 DKK pr. Kg. of saturated fat**

	10. percentile	Median	90. percentile	Mean
Butter	0.96	1.16	1.44	1.20
Mixed butter	0.98	1.13	1.54	1.23
Hard margarine	1.46	1.96	3.00	2.12
Spread marg.	0.48	0.91	1.34	0.92
Minarine	0.25	0.30	0.35	0.30

Household  $h$ 's specific price change is based on household  $h$ 's average purchase prices. It is assumed that all of the tax is put on the consumer.

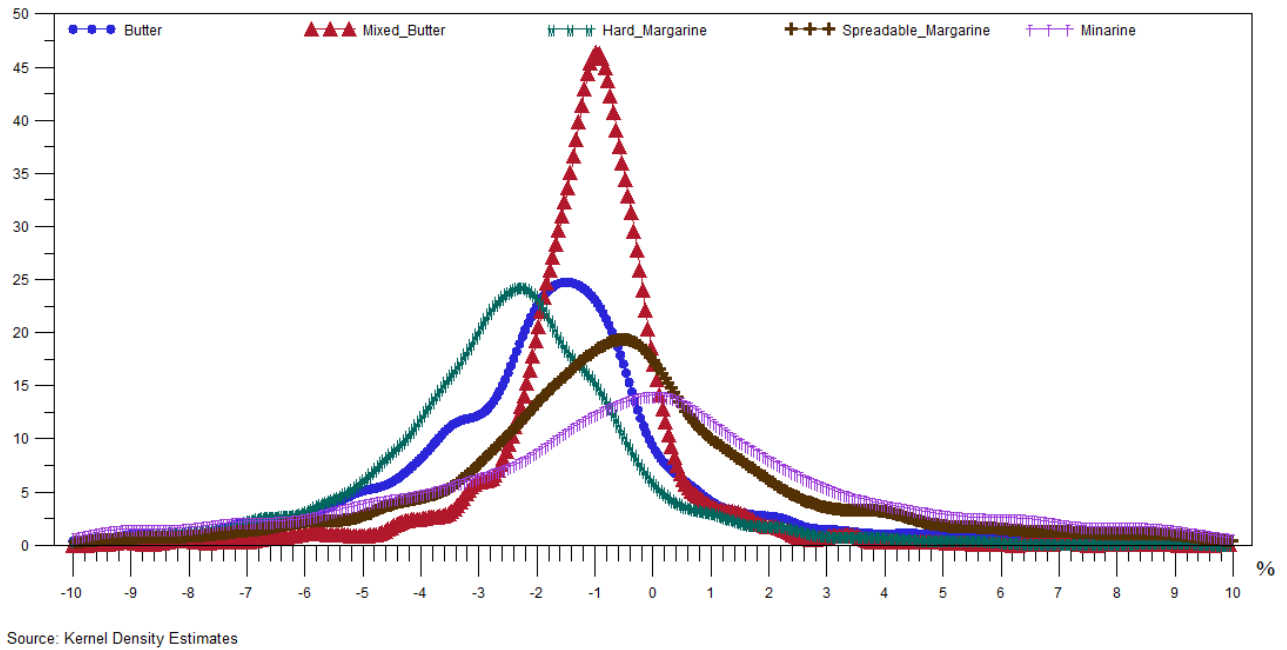
It is important to note that although *butter* will increase most in price (in absolute terms) because of its high saturated fat content, its relative price change (which is used for simulation) will not be the largest because the initial price of *butter* is higher than the other goods.. Even within the same class of goods, e.g. butter, differing (average) relative price changes between households can be observed because they buy butter at different prices because of differing characteristics such as brand etc. Hence, the percentage change in prices induced by the tax reform is household specific.

Especially the *hard margarine* exhibits a relatively large percentage price change, due to high saturated fat content and a low initial price.

Under the assumption that the total effect of the tax equals the sum of its partial effects, the total percent change of good  $i$ ,  $i = 1, \dots, 5$ , for household  $h$  can be calculated as:

in which  $\epsilon_{ij}^k$  is the price elasticity of household  $k$  for good  $i$  with respect to the price of good  $j$  defined in equation (4.1);  $\Delta p_j^h$  is the household specific percentage price change of good  $j$  for household  $h$ ,  $x_j^h$  is the after-tax purchases of good  $j$  for household  $h$  and  $x_j^h$  is the before-tax purchases of good  $j$  for household  $h$ . The distributions of the relative consumption changes for these 5 goods are shown in figure 3.

**Figure 3 Relative demand change ( %) from a tax of One DKK per Kg of saturated fat**



The households' likely consumption response to a tax of one DKK is quite heterogeneous. A sizeable proportion of the households have a percentage change which is double, triple or even quadruple as much as other households. Some households decrease their consumption of a given good, while other households increase their intake of the same good. This heterogeneity within population price responsiveness underlines the importance/benefits of the adopted approach.

Due to the largest share of saturated fat in *butter*, one might expect *butter* to be the good that households reduce their consumption of the most in response to the tax on saturated fat. But this is not the case; the distribution of change in *hard margarine* stochastically dominates those of all of the other goods. This is the case because *hard margarine* is relatively cheap to start with and the tax implies the largest relative price changes for *hard margarine* even though its content of saturated fat is lower than that of *butter*.

An understandable reason for the strong price response is the relatively close substitutability between the goods. Thus, it is likely that the unhealthiest alternatives are sacrificed for the benefit of the healthier choices for a large share of the population, i.e. it is quite clear that the distributions in the net decrease in *hard margarine* and *butter* stochastically dominate the corresponding distributions for *spreadable margarine* and *minarine*. The consumption of

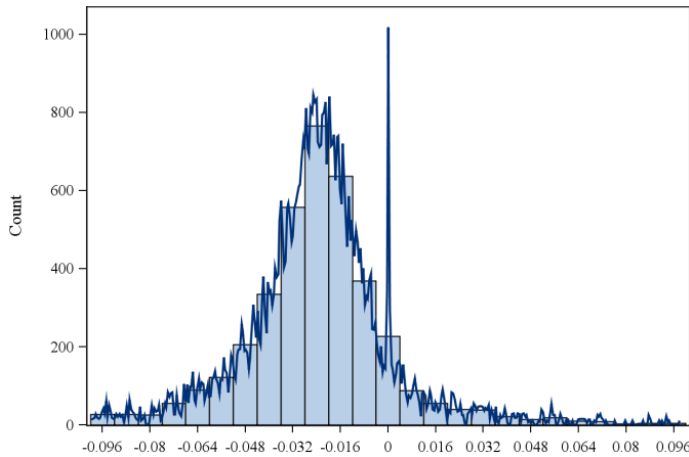
*mixed butter* is also mainly decreased and approximately half of the households increase their intake of *minarine*, even though its price increases. These households are behaving as if *butter*, *mixed butter* and/or *hard margarine* are substitutes to *minarine*, and accordingly they substitute in favour of *minarine* as a consequence of the new price system implying a higher, but relatively favourable, *minarine* price.

This may seem promising from a health promotion perspective. However, here we should keep in mind that not all households purchased *minarine*, actually *minarine* is the least frequent buy of the five alternatives (see Table A1), and it is only half of the subpopulation who do purchase *minarine* that we can expect to increase their consumption of *minarine*.

For comparison, an alternative (VAT) simulation was conducted, calculating the demand effects of increasing the prices on all the considered fat products by the same percentage. Such a tax change yields decreased demand for all goods in the order (with the largest decreases first): Butter, Minarine, spreadable Margarine, Mixed Butter and Hard Margarine. The distribution of the change in butter stochastically dominates the (distribution of the) changes in the other goods when all prices increase by one percent. Whereas the relatively strong reduction in butter consumption is consistent with the aim of reducing saturated fat intake, the healthier alternatives minarine and spreadable margarine would decrease more than the unhealthier alternatives mixed butter and hard margarine. Although some of the alternatives considered in this analysis might be substituted with healthier alternatives such as edible oils, it is even more likely that minarine and spreadable margarine will be substituted by mixed butter and hard margarine. A VAT on the considered fat products is likely to be less successful than a tax on SFAs in manipulating the Danes' diet towards products with less SFAs and might even be worse than no tax introduction, because its ability to motivate households to substitute away from hard margarine in favour of alternatives with less saturated fat, might be limited. This is in contrast with the saturated fat tax, where the households' relatively insensitive response to price changes regarding hard margarine demand is not a matter of concern because hard margarine has a relatively high share of saturated fat and has a low average price.

It should be noted that each distribution in figure 3 is based solely on households having made purchases of the corresponding good. Hence, the number of households that did not make any purchase of a considered good (thus having a tax induced change in demand of zero) are not included in these distributions. If the households that did not make any purchases for a considered good are included, there would be a giant spike around zero for each good (see, for example, figure 4 for hard margarine below). This is especially the case for minarine, which the majority of households did not purchase. The distribution of demand change in Figure 4 is clearly under smoothed, but this is necessary to obtain the spike around zero which indicates that a substantial share of the population never buy hard margarine.

**Figure 4 Relative demand changes for Hard Margarine from a Tax of One DKK per Kg of saturated fat**



## 5 Discussion and conclusion

In the paper, a method for imputing missing prices based on the discrete choice framework, as a basis for estimating demand systems and price elasticities for butter and margarine products at the household level, and for simulating effects of alternative tax schemes, has been introduced.

The developed methodology demonstrated in this paper can be useful in solving other policy relevant questions. For instance, a tax ministry might want to have an estimate of the tax revenue for a given tax rate. Using household specific price, quantity and price elasticities in calculating the demand response would probably provide more accurate estimates of the effect of a tax on total change in demand, compared to a more conventional methodology. Let  $\theta_i$  be the share of households that change their demand for good  $i$  with  $\theta_i$  in response to the tax, then the total change in demand for good  $i$  is  $\theta_i \Delta q_i$ . Using the methodology of this paper, the effect could be estimated by:

using similar notation as in equation (4.3). However, using a more conventional methodology that uses average estimates would mean that the total change in demand for good  $i$  of a tax would be estimated by

$$\Delta q_i$$

The estimate in expression (5.1) of the total change is likely to be more accurate, especially considering the non-symmetrical shape of the estimated price elasticities, when using the methodology in this paper compared to a more conventional approach based on average estimates.

Another example is that policy makers might want to know what the consequences of a tax will be for an industry. Again, the methodology of this paper can be applied to give more accurate estimates of the expected changes in demand.

In the simulation experiment, it is implicitly assumed that the tax is fully transmitted to prices, i.e. a 1:1 transmission for all prices and all retailers is assumed to respond identically, but this might not be the case. There is evidence that

suggests that a tax will not be fully transmitted to prices. For instance, an incomplete transmission of coffee bean prices to consumer prices in the Netherlands has been observed (Bettendorf & Verboven, 2000). Also, Weiss (1995) finds that cost changes are less fully transmitted into prices in highly concentrated industries in Austrian manufacturing. The qualitative results will still hold if a tax is not fully transmitted to prices, but it is transmitted in a uniform way over products. The market structure of retail in Denmark is concentrated as it can be characterised as a duopoly (Vorley, 2007) and it might be reasonable to assume that retailers would use their market power and information on the households' consumption behaviour to adjust consumer prices in a profit optimising way in response to input price changes implied by the tax and thus prices would not necessarily increase uniformly. This paper, and the papers cited in the introduction, does not model strategic pricing behaviour in retailing, but this could be a subject for further studies. However, modelling strategic pricing behaviour might not be an easy task, since direct empirical evidence on the relative importance of retailer upstream and downstream market power on consumer prices is not available according to Weiss & Wittkopp (2005). A paper that seeks to simulate the effect of a tax introduction that takes the strategic pricing behaviour in retailing into account is Griffith *et al.* (2010).

The majority of the sample, and accordingly the majority of the Danish population, have elastic demand for all of the considered goods, but the share of inelastic demand behaviour is larger than 25 % for all the considered goods. This paper considers the effects of taxing saturated fat in butter and margarine products on the consumption of these goods. The analysis reveals that a likely outcome of a saturated fat tax for the majority of the population is a reduction in purchases of Hard Margarine and butter products. These will, to some extent, be substituted with healthier alternatives such as spreadable margarine and minarine. It is therefore expected that this tax will improve the Danes' fat intake profile, i.e. the total intake of fat might drop slightly, and saturated fat will comprise a smaller proportion of total fat intake.

The effects of a tax targeted on saturated fat seem to be superior to a VAT in that a likely outcome of the former is improved diets. This is a much less likely outcome in the case of the VAT scenario.

It is recognised that the analysis in this paper does not provide the full picture of the effect of the tax, but as these products constitute an important share of the SFA intake, the study is considered to provide some indications of the effects of the tax proposal. However, other alternatives to the goods considered in the analysis do exist. Edible oils are clearly substitutes for the considered goods in this paper. It seems reasonable that some of the likely reductions in purchases of butter, mixed butter and hard margarine will be substituted by purchases of healthier vegetable oils, thus improving diets in Denmark.

While the majority of the population is expected to change their behaviour in a way that improves their diet (and consequently reachable through a fat tax), there is also a part of the population which will not, e.g. some households are likely to increase their intake of butter as they may find that hard margarine has become too expensive relative to butter. A price policy, thus, has the potential to trap individual households in inappropriate behaviour that can counteract the purpose of improving the Danes' diet, something which must be evaluated carefully.

The households' relative percentage price change induced by the tax on saturated fat depends on their current choices,

reflecting the preferences of the households. This was shown to exhibit large heterogeneity. This together with great heterogeneity within population price responsiveness (measured by elasticities) makes it important that prices used in evaluating the tax reform are constructed in the best possible way, which underlines the benefits of the adopted approach. These heterogeneous price effects should be considered when developing policy that is designed to improve diets and carefully evaluated to minimise the number of individual households trapped by policy into regressive behaviour. Especially taxes similar to the tax on saturated fat considered in this paper might give unexpected policy consequences, as the relative percentage change in prices is household specific depending on preferences.

Although, the European Food Safety Authority recommends that SFA intake should be as low as possible within the context of a nutritionally adequate diet, recent meta-analysis based studies (Mozaffarian *et al.*, 2010) and an expert panel review of the evidence of SFA as a risk factor (Astrup *et al.*, 2011) gives evidence towards the hypothesis that it is more important what the saturated fatty acids is substituted with than merely reducing the SFA intake. Astrup *et al.* (2011) finds consistent evidence that the risk of CHD is reduced when SFAs are replaced with poly unsaturated fatty acids (PUFAs), but no clear benefit of substituting carbohydrates for SFAs has been shown, although there might be a benefit if the carbohydrate is unrefined and has a low glycemic index, and insufficient evidence exists to judge the effect on CHD risk of replacing SFAs with monounsaturated fats (MUFAs). Astrup *et al.* (2011) further argues that the effect of particular foods on CHD cannot be predicted solely by their content of total SFAs because individual SFAs may have different cardiovascular effects and major SFA food sources contain other constituents that could influence CHD risk. Astrup *et al.* (2011) finally concludes that research is needed to compare specific foods with appropriate alternatives. Policy makers are interested in improving people's diets through pricing instruments. Thus, the learning outcome of this literature review may be that it might not be a good idea to introduce a uniform tax on SFAs in foods. Rather it would be better to introduce a tax on foods containing SFAs where better alternatives exist, e.g. a SFAs tax on butter could be beneficial, because a healthier alternative, rapeseed oil, exist.

It is possible that a number of households are not price sensitive, or that unexpected and unfortunate policy consequences will arise that increase some households' SFAs to PUFAs ratio. This suggests that pricing policies alone cannot solve diet related problems such as obesity and coronary heart disease (CHD), only aid in improving the obesity and CHD prevalence rates. There are other policy instruments that can help reduce obesity and CHD prevalence. For instance, a ban on trans fatty acids (TFAs) in foods in Denmark has reduced the Danes' intake of TFAs, which is likely to contribute in a positive way to the prevalence of CHD (L'Abbe *et al.*, 2009). A ban on SFAs in food would have an effect on individuals who do not respond to price increases, but would be unacceptable (see Astrup *et al.*, 2011). However, other policies, such as information campaigns and policies that aim to increase physical activity could be



beneficial. For instance, information campaigns have affected the Danes' intake of butter in a positive way (Holm *et al.*, 2002). So, individuals who do not respond to price changes might respond to information. However, individuals who are price insensitive are more likely to have higher utility from consuming the foods and consequently they would be more difficult to influence through information. These individuals might, however, be responsive to policies that aim to increase their physical activity level which could reduce the risk of them becoming obese or suffering from CHD. Thus, a strategy that involves a larger spectrum of policies is likely to be more successful and, therefore, such a strategy is recommended.

## References

Astrup, A.; Dyerberg J; Elwood, P.; Hermansen, K.; Hu, F. B.; Jakobsen, M. U.; Kok, F. J.; Krauss, R. M.; Lecerf, J. M.; LeGrand, P.; Nestel, P.; Risérus, U.; Sanders, T.; Sinclair, A.; Stender, S.; Tholstrup, T. and Willett, W. The role of reducing intakes of saturated fat in the prevention of cardiovascular disease: where does the evidence stand in 2010? *Am J Clin Nutr* 2011 ajcn.004622; First published online January 26, 2011. doi:10.3945/ajcn.110.004622.

Brownell, Kelly D., and Frieden, Thomas R.(2009). Ounces of Prevention — The Public Policy Case for Taxes on Sugared Beverages. *The NEW ENGLAND JOURNAL of MEDICINE*. April 30, 2009. Available at <http://content.nejm.org/cgi/reprint/NEJMp0902392.pdf>. Accessed on 15 Marts 2010.

Cox, T. L., and Wohlgenant, M. K. (1986). Prices and quality effects in cross-sectional demand analysis. *American Journal of Agricultural Economics*, 68, 909-919.

Deaton, A., and Muellbauer, J. (1980a), *Economics and Consumer Behavior*, Cambridge, England: Cambridge University Press.

Deaton, A., and Muellbauer, J. (1980b), "An Almost Ideal Demand System," *American Economic Review*, 70, 312-336.

Dhebbi B., Gil J.M. & Angulo A.M. (2007) Nutrient Effects in Consumer Demand Systems: Evidence from Panel Data, *Food Economics*, vol. 4, pp. 89-102.

Edgerton D.L. (1997) Weak separability and the estimation of elasticities in multistage demand systems. *American Journal of Agricultural Economics*, vol. 79, pp. 62-79.

European Commission (2007). WHITE PAPER ON A Strategy for Europe on Nutrition, Overweight and Obesity related health issues. Brussels, 30.5.2007. Available at [http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/documents/nutrition\\_wp\\_en.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/nutrition_wp_en.pdf)

Accessed on 15 Marts 2010.

European Food Safety Authority (2009). Scientific Opinion of the Panel of Dietetic Products, Nutrition and

- Allergies on a request from the European Commission related to dietary reference values for fat. The EFSA Journal (200x) xxx, 1-97. Available at:  
[http://www.efsa.europa.eu/en/calls/documents/nda\\_op\\_drv\\_fat\\_draft\\_en\\_released%20for%20consultation.pdf](http://www.efsa.europa.eu/en/calls/documents/nda_op_drv_fat_draft_en_released%20for%20consultation.pdf). Accessed:11.3.2009:
- Forebyggelseskommissionen (2009): Vi kan leve længere og sundere – Forebyggelseskommissionens anbefalinger til en styrket forebyggende indsats, Betænkning 1506, April 2009.
- GfK ConsumerScan(2010). Available at <http://www.gfkcs.com/index.php?cat=1>. Accessed on 8 April 2010.
- Greene, W.H. (2003): *Econometric Analysis*, 5<sup>th</sup> ed., pp. 913f. Prentice-Hall, New Jersey.
- G.W. Gustavsen, D. Jolliffe og K. Rickertsen (2008). "Censored Quantile Regression and Purchases of Ice Cream," *Food Economics*, 5: 152-163.
- Ghysels, E. and D. R. Osborn, 2001. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, Cambridge.
- Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(486): 1015-1026.
- Huang K.S. & Lin B.H. (2000) Estimation of Food Demand and Nutrient Elasticities from Household Survey Data, Technical Bulletin no. 1887, Economic Research Service, US Department of Agriculture, Washington DC, 30 pp.
- Jensen J.D. & Smed S. (2007) Cost-effective design of economic instruments in nutrition policy, *International Journal of Behavioural Nutrition and Physical Activity* (<http://www.ijbnpa.org/content/4/1/10>)
- Koc A., Dolekoblü T. & Erdal Ertuk Y. (2001) Household Vegetable Oil and Butter Demand in Turkey: A Three Stage Budgeting and AIDS with Adjusted Unit Value, paper presented at the 71<sup>st</sup> EAAE Seminar: The Food Consumer in the Early 21<sup>st</sup> Century, Zaragoza (Spain), April 19-20, 2001
- McFadden D. (1974). *Conditional Logit Analysis of Qualitative Choice Behaviour*, in P. Zarembka (ed.), *Frontiers of Econometrics*, Academic Press.
- Mozaffarian D., Micha R., Wallace S.(2010) Effects on Coronary Heart Disease of Increasing Polyunsaturated Fat in Place of Saturated Fat: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *PLOS Medicine*. Volume 7. Issue 3, pp 1-10.

Nichele V. (2003) Health Information and Food Demand in France, chapter 8 in Chern W.S. & Rickertsen K. (eds.) Health, Nutrition and Food Demand, CAB International

Pedersen A. N., Fagt S., Groth M. V.(2010). Danskernes kostvaner 2003 – 2008. DTU Fødevareinstituttet. January 2010. ISBN: 978-87-92158-67-3.

Platform Charter(2005). EU Platform on Diet, Physical Activity and Health. DIET, PHYSICAL ACTIVITY AND HEALTH - A EUROPEAN PLATFORM FOR ACTION. 15 March 2005. Available at [http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/platform/docs/platform\\_charter.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/platform/docs/platform_charter.pdf). Accessed on 15 Marts 2010.

Racine, J. S. (2002). Index-free density-based multinomial choice. In Ullah, A., Wan, A., and Chaturvedi, A., editors, Handbook of Applied Econometrics and Statistical Inference, pages 115-142. Marcel Dekker.

Rickertsen K. (1998) The Demand for Food and Beverages in Norway. Agricultural Economics, vol. 18, pp. 89-100.

Smed S., Jensen J.D. & Denver S. (2007) Socio-economic characteristics and the effect of taxation as a health policy instrument, Food Policy, vol. 32, pp. 624-639

The Danish Ministry of Taxation (2009,I). Lovforslag om afgift på mættet fedt, 18.08.2009. Available at <http://www.skm.dk/public/dokumenter/lovstof/Lovforslag%20maettet%20fedt.pdf>. Accessed on 15.3.2010.

The Danish Ministry of Taxation (2009,II). Danish Tax Reform 2010 - Paper to the OECD WP 2 meeting November 2009 by The Danish Ministry of Taxation. Available at [http://www.skm.dk/public/dokumenter/engelsk/Danish%20Tax%20Reform\\_2010.pdf](http://www.skm.dk/public/dokumenter/engelsk/Danish%20Tax%20Reform_2010.pdf). Accessed 15.3.2010.

Yen S.T. & Chern W.S. (1992) Flexible Demand Systems with Serially Correlated Errors: Fats and Oil Consumption in the United States, American Journal of Agricultural Economics, vol. 74(3), pp. 689-696

WHO (2003, a). Diet, nutrition and the prevention of chronic diseases: report of the joint WHO/FAO expert consultation.

WHO (2003, b). Obesity and Overweight. Global Strategy on Diet, Physical Activity and Health. Available at <http://www.who.int/dietphysicalactivity/media/en/gsf Obesity.pdf>. Accessed on 15 Marts 2010.

WHO (2004). Global Strategy on Diet, Physical Activity and Health. Available at [http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy\\_english\\_web.pdf](http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf). Accessed on 15 Marts

2010.

WHO (2006). Fact sheet N°311, September 2006. Available at:

<http://www.who.int/mediacentre/factsheets/fs311/en/index.html>. Accessed on 15 Marts 2010.

# Paper 5: Determinants of development in inequality in fruit and vegetable consumption in Denmark

---

## **Abstract**

**Purpose** - Inequality in consumption of fruit and vegetables (FAV) in Denmark has increased in a period of extensive information campaigns, which had the purpose of increasing the consumption of FAV. The paper attempts to identify the reasons for this development.

**Design/methodology/approach** – Quantile regression is used on a rich data set from the market research institute GfK Denmark for the analysis of the development in FAV consumption. The determinants of the development in FAV consumption are investigated by use of a new methodology that enables comparisons of data samples with different covariate distributions.

**Findings** – The study shows a negative relationship between the intensity of consumption and price sensitivity. Increased inequality in consumption over time can partly be attributed to this negative relationship coupled with increases in prices, but to a greater extent so that uneducated groups with a low consumption and low income groups are falling behind the more educated and financially well off segment of the population.

**Practical implications** - The findings may be utilised to target specific sub-populations when designing policies aimed at increasing fruit and vegetable consumption.

**Research implications** – These findings highlight the advantages of quantitative methodologies that compare developments in samples and adjust for differences in covariate distributions.

**Originality/Value** – The paper utilises new econometric methods to analyse the determinants of the development in inequality in FAV consumption.

**Keywords** Fruit and Vegetable consumption, Information Campaign, Designing diet policies.

## 1. Introduction

A WHO/FAO<sup>27</sup> expert consultation report on diet, nutrition and the prevention of chronic diseases recommends an intake of a minimum of 400 g of fruits and vegetables (FAV) per day for the prevention of chronic diseases such as heart disease, cancer, diabetes and obesity WHO/FAO(2003). The report states that there is convincing evidence that fruit and vegetables decrease the risk of obesity. Eating a variety of vegetables and fruit ensures an adequate intake of dietary fibres and can help displace foods which are high in saturated fats and sugar. Low FAV intake is present in the top 8 of leading risk factor causes of death in middle and high income countries (WHO, 2004).

Also, the U.S. Department of Health (2004) stresses the importance of communicating nutrition messages including the National Cancer Institute's 5-A-Day for Better Health Program to increase FAV consumption in order to address the obesity epidemic. The European Commission (2005) is interested in identifying measures that could contribute towards improving the attractiveness, availability, accessibility and affordability of fruit and vegetables. All these institutions have recommendations and policies regarding the consumption of FAV. A prominent example of a policy at the supranational level is the reform within the Common Agricultural Policy of the Common Market Organisation (CMO) for fruit and vegetables, which is aimed at promoting FAV consumption within specific settings, such as schools. Part of the reform will promote children's consumption of FAV through proposals to allow surplus production to be distributed to educational institutions, and children's holiday centres. Additionally, a school fruit scheme offering greater affordability of FAV to encourage consumption is part of the reform of the CMO for fruit and vegetables (European Commission, 2007).

The EU commission has recommended public-private partnerships towards consumer education regarding healthy food choices (European Commission, 2008). Many countries have adopted government supported campaigns aimed at increasing FAV consumption: “5 a day” is such an example and it has been adopted in various English speaking countries, notably the United States (Havas *et al.*, 1995) as well as England, Wales and Scotland. In Denmark, a similar campaign is the “6 a day” campaign, which was agreed upon by central nutrition educators in September 1998 (Ministry of Food, Agriculture and Fisheries, 2001). It is a public-private partnership with representatives from government agencies, non-governmental health organisations, consumer organisations and industrial sectors including the FAV industry and the meat industry<sup>28</sup>. The primary goal of the campaign is to increase the consumption of FAV among Danish consumers (European Commission, 2008).

Governments initiate information campaigns with the expectation that they will affect the behaviour of the population. In a cost benefit context, it is important to investigate the effectiveness of such campaigns with respect to health achievements. But governments are also interested in the inequality in the population because of the stability of society and moral concerns. In Denmark, a reduction in inequality in health is a goal of the current government and was also a goal of the previous left-wing government (Government, 2002), whilst it is also a goal of the 6-a-day campaign (Ministry of Food, Agriculture and Fisheries, 2001, p.5)<sup>29</sup>. Therefore, monitoring factors which influence health is

---

<sup>27</sup> World Health Organization/Food and Agricultural Organization.

<sup>28</sup> The Danish Meat Association contribute to 6-a-day with knowledge about meat as part of a meal, culinary cooking, recipe development, etc. and by providing network contacts such as dieticians, teachers of home economics, food writers etc.

<sup>29</sup> The policy objective of the UK “five-a-day” campaign also includes a reduction in health inequalities (Capacci,

important in a cost benefit context, but also when evaluating whether government initiatives are increasing equality or inequality. There is perhaps reason to fear that information campaigns will increase inequality in health if only subgroups of the population respond to the campaign message. One might imagine that the educated segment of the population would be more likely to respond to a campaign message if decoding it requires formal schooling; if education helps develop strategies to change behaviour and if persons who are more responsive to authorities, self-select into further education. Finally, the educated segment of the population may be more likely to respond to the campaign message if the information is focused towards this particular segment of the population. Actually, this is the case for the 6-a-day information campaign, which, according to the campaign strategy, should focus on the segment of the population who are educated and “ready to change”. The rationale is that these groups are motivated to comply with nutritional advice and are expected to act as opinion leaders (Ministry of Food, Agriculture and Fisheries, 2001). Also, if an information campaign increases average demands it might drive up prices, which can harm low consuming groups if they are more price sensitive. Validating this is obviously an empirical question.

The “6 a day” campaign has occurred at the same time as an increased intensity in the information flow of health related issues. This suggests that the discrepancy in the intake of FAV between uneducated and educated subgroups should increase if the above hypothesis is correct. This paper analyses the development in the purchases of FAV. While it is not possible to evaluate the effect of information campaigns such as “six a day” on behaviour directly (differentiating the effect of other information sources and drivers), it is possible to analyse whether the discrepancy in purchases of FAV imposed by educational differences, is increasing or not. If true, it would support the hypothesis that the information has worked, but only on a subset of the population.

This study attempts to address this issue by using a dataset on actual purchases, a linear quantile regression model and a recently developed non-parametric approach to find a correctly specified model for our examination of the determinants of FAV consumption. The study thus provides timely information on the important determinants of FAV consumption in Denmark. Also, in previous studies based on record and recall, the focus has been on the reported quantity consumed, while this study analyses the actual purchases. Obviously it is the food consumed that is of interest in an obesity context, but actual shopping behaviour is also of interest for many reasons<sup>30</sup>. Firstly, the participants in the study’s survey are less prone to cheap talk or a social desirability bias compared to recall and record studies. Cheap talk can be a problem in record and recall studies because participants know what the most socially acceptable behaviour is and they incur no cost if they deviate from the truth<sup>31</sup>. Secondly, prices are not incorporated in record and recall studies, but it is important to determine whether price is a critical determinant of FAV consumption, since it is a directly controllable policy instrument. Finally, it is important that an evaluation of the 6-a-day campaign does not merely measure the change in the sample intake of FAV<sup>32</sup>, but also accounts for changes in the composition of distribution of

---

2010).

<sup>30</sup> Studies show significant positive correlations between household-level food purchases and dietary intake by individuals (Eyles *et al.* (2010) and Ransley *et al.* (2001) ).

<sup>31</sup> The primary objective of these studies is to analyse dietary behaviour in contrast to the data analysed in this paper which, first and foremost, have been used by supermarket chains to set prices and similar marketing activities. Studies show that purchase data have less response bias relative to self-reported dietary measures (Hebert *et al.* (2008), Kristal *et al.* (1998)).

<sup>32</sup> One evaluation of the 6-a-day campaign focused on measuring the change in sample intake of FAV and the amount

socio-demographic variables such as education and income and accounts for the development in prices. An evaluation that does not account for such factors can give rise to misleading results (Gordon *et al.*, 2006).

This paper is organised as follows. Section 2 discusses the data source and data description and briefly introduces testing linear mean and quantile regression models, as well as the quantile regression model. In addition, the procedure for comparing the sample distribution is presented. Section 3 discusses the estimation results based on both a basic parametric linear model and the model obtained by virtue of applying the consistent model specification test discussed in Section 2. Conclusions are presented in Section 4.

## 2. Data description and econometric models

In this section, we provide an overview of the household data used in the present analysis. We then briefly introduce non-parametric consistent specification tests. Finally, the methodology for comparing expenditure distributions is presented.

Several researchers and institutions have monitored and investigated FAV expenditure/consumption in different countries and based on different data sources<sup>33</sup>. For a recent evaluation of the UK five-a-day program see Capacci *et al.* (2010) and for a systematic review of the literature on interventions designed to increase adult FAV consumption see Pomerleau *et al.* (2005).

The only information available on Danish households' consumption of Fruit and Vegetables (FAV) prior to 1995 comes from the National Statistics Account, Statistics Denmark. Beside this, the development in the consumption of FAV has been analysed since 1995 in two studies (see Fagt (2008), Haraldsdóttir *et al.* (2005), both of which utilise record and recall surveys.

Fagt *et al.* (2008), found that the average consumption increased when comparing the periods 2000-2002 to 2003-2006 (see Table 1).

---

of a sample that had knowledge about the campaign (Danish Cancer Society, 2006). A substantial rise in awareness has been observed in Denmark (Danish Cancer Society, 2006) and in the UK (Bremmer *et al.*, 2006) as a result of the campaigns, which is also a common finding in the evaluation of social marketing initiatives (Gordon *et al.*, 2006).

<sup>33</sup> Forward looking interesting new possibilities are becoming available through the development of new initiatives to monitor the development of indicators on the health status of European citizens including the frequency of consumption of FAV and obesity. The European Commission in particular has developed a European Health Interview Survey (EHIS), which has been operational since 2007 and allows the collection of measures through the European Statistical System (Eurostat, 2010). The release of the EHIS anonymous micro data to the research community is scheduled on June 2011 (Eurostat, 2010). Furthermore, a two-year planning and pilot phase of a European Health Examination Survey (EHES) started in 2009 to obtain objective information on a range of measures including BMI, and Blood samples like HDL cholesterol in a randomly selected population (EHES, 2009).



Table 1 Consumption of FAV by individuals aged 11-75 in grams separated by quantile and period

Quantile	4	17	18	37	40	57	60	73	76	84	86
2000-2002	100	-	200	-	300	-	400	-	500	-	600
2003-2006	100	200	-	300	-	400	-	500	-	600	-

Source: Weekly recorded diaries; (Adapted from Fagt *et al.*, 2008) Note: ‘-’ indicates that no information is available.

The reason for the lack of information in some cells is that we have transformed the data in order to better compare it with our results on Gfk data.

Other data sources also support an increase in the average consumption in a similar timeframe, e.g. the Danish Cancer Society (2006) reports sample average of 2.82 servings of FAV in 1999 and 3.46 servings of FAV in 2003, and according to National Accounts, the private consumption of FAV except potatoes was 8.683 billion DKK in 1999 and 10.519 billion DKK in 2004 (current prices) (National Accounts 2007 and National Accounts 2005), a development which has occurred during a period of moderate price increases (Table 2)<sup>34</sup>.

Table 2 Percentage annual increase in prices and price index of FAV except potatoes by year

Year	1999	2000	2001	2002	2003	2004	2005
% annual price change		-0.5	6.4	1.9	4.1	-2.8	-0.3
Price Index	100	99.5	105.87	107.88	112.3	109.16	108.83

(Source: National Accounts, 2007:National Accounts, 2005 and own calculations).

When analysing healthy eating behaviour, specific target groups such as individuals who don’t eat much, are of more interest than groups, which have an average level of consumption. This explains why quantile regression has become increasingly popular when analysing eating behaviour. Gustavsen & Rickertsen(2006) use a censored linear quantile model, which takes the zero purchase in the considered survey periods into account. Stewart *et al.* (2003) also use a censored linear quantile model and a test for stochastic dominance is performed to determine whether high income households stochastically dominates low income households. Both of these studies use data from Consumer Expenditure Surveys collected by a national statistics bureau and use a linear specification of the quantile regression model, but do not test if the model is mis-specified. There are no Danish studies on the determinants of FAV consumption that use quantile regression.

## 2.1. Data description

The present study uses household panel data based on a “nation-wide representative sample,” which “is selected according to geographic and demographic criteria based on Central Statistical Bureau information” (GfK Consumer Tracking, 2010). GfK panel members produce weekly diaries in which they report household purchases regarding quantity, price, expenditure, and other relevant product features etc. on a detailed level. The total sample sizes are 1306 households in 1999 and 1927 in 2004.

<sup>34</sup> Table 2 indicates that the price increase from 1999 to 2004 was 9.2 %.

The analysis of the determinants of FAV consumption uses reported purchases of FAV composed of 32 fruit categories and 41 categories of vegetables. Some of the FAV items are measured in grams e.g. 200 grams of raspberries, whilst others are referred to according to the number of items purchased e.g. ten apples. To account for this conversion, factors<sup>35</sup> were constructed that allowed all FAV items to be measured in grams. For example, the average weight of one apple is 139 grams and so a purchase of ten apples yields a purchase of 1390 grams of fruit. The purchases of FAV in grams entered in the weekly diaries are aggregated to give the yearly purchase of FAV in grams for 1999 and 2004. The dependent variable in the regression analysis is the log of the household's yearly FAV consumption in grams<sup>36</sup>.

In addition, Gfk panel members completed an annual questionnaire on socio-demographic characteristics from which the following explanatory variables are constructed: (i) Household income (I) divided into three intervals: (1) 0-200,000 Danish Kroner, (2) 200,000 – 450,000 Danish Kroner and (3) at least 450,000 Danish Kroner; (ii) household size (HS); (iii) education level of the main shopper (E) with five categories: (1) middle-level specialised training, (2) 2-year college, (3) bachelor's degree, (4) Master's degree, (9) uneducated; (iv) age (A) of the main shopper (diary keeper); (v) a gender dummy (S) of the main shopper. Finally, a sixth explanatory variable - a price index P - is constructed from the weekly diaries. The explanatory variables are represented by , .

#### Construction of price index

The price index (Stone index) is constructed such that prices on FAV items with higher expenditure shares are given higher weights thus:

in which  $P_{H,t}$  is the price index of household  $H$  in sample period  $t$ ;  $E_{H,t}$  is the total expenditure on food item  $k$  for household  $H$  in sample period  $t$ ;  $E_{FAV,t}$  is the total expenditure on FAV for household  $H$  in sample period  $t$ ;  $P_{k,t}$  is the average price per gram of food item  $k$  for household  $H$  in sample period  $t$ ;  $Q_{k,t}$  is household  $H$ 's  $k$ 'th purchase of food item  $k$  in sample period  $t$  measured in grams.

## **2.2. Testing parametric mean and quantile regression models**

To decide whether a parametric model is statistically significant, we apply a recently developed model specification test by Hsio, Li, and Racine (2007) when dealing with a parametric mean regression, and the test described in Racine (2006) when testing a parametric quantile regression model. For a detailed treatment of the applied tests see Hsio, Li, and Racine (2007) for the mean regression test and Zheng (1998) and Racine (2006) for the quantile regression test.

<sup>35</sup> Most of the conversion factors used are available in *Typical weights for food items* (Andersen *et al.*, 1996).

<sup>36</sup> Consumption should be understood as the household's yearly purchase. Studies show a significant positive correlation between household-level food purchases and the dietary intake of individuals (Eyles *et al.* (2010) and Ransley *et al.* (2001) ).

The null hypothesis of these tests is that the parametric model is in fact a correct specification, while the alternative hypothesis is that the parametric model is mis-specified with the nonparametric model being the correct specification. Roughly speaking, the test is based on the difference between a parametric fit and a nonparametric fit using the cross-validation method to select smoothing parameters.

## 2.3 Quantile regression

Simple mean regression cannot satisfactorily fit the data in many cases if the distributions of the error term are non-Gaussian and asymmetrical and/or have a fat tail, or if the dependent variable includes outliers, which often happens in household survey data. Under these circumstances, the conditional mean estimator can be sensitive to the presence of outliers and can thus be misleading. The quantile regression introduced by Koenker and Basset (1978) is an alternative method to provide estimates of a dependent variable, which corresponds to various quantile values of the explanatory variables so that a more comprehensive picture of the conditional distribution of a dependent variable can be obtained. More importantly, the quantile process provides a more comprehensive picture of the data, which is of interest when analysing food demand behaviour.

### 2.3.1 Parametric Regression

For a random variable  $Y$  with probability distribution function  $F_Y(\cdot)$ , the  $\tau$ th quantile of  $Y$  is defined as the inverse function  $F_Y^{-1}(\tau)$ , where  $\tau \in (0, 1)$ .

For a random sample  $Y_1, \dots, Y_n$  of  $Y$ ,  $X$ , quantile regression estimates the linear conditional quantile function:

by solving:

for any quantile  $\tau$ . The quantity  $\hat{\beta}_\tau$  is called the  $\tau$ th regression quantile. The set of regression quantiles  $\{\hat{\beta}_\tau\}_{\tau \in (0, 1)}$  is referred to as the quantile process. The problem of finding  $\hat{\beta}_\tau$  is solved as a linear programming problem.

### 2.3.2 Modelling distributions

Following the problem formulation, the main interest is the change in purchases of fruit and vegetables during the period 1999-2004. Time is the treatment, so the households in the 2004 sample can be considered participants, whilst

the households in the 1999 sample can be considered non-participants. In treatment evaluation, only the participation outcomes for the participants and the non-participation outcomes for the non-participants can be observed, but not the outcomes those treated would have realised had they not been treated and vice versa, i.e. their counterfactual outcomes. However, when estimating treatment effects, it is necessary to compare the counterfactual outcomes with the realised outcomes. Inferring how much FAV the households in the 2004 sample (participants) would have consumed had they not been experiencing the effect of the 6-a-day campaign, i.e. the counterfactual outcomes for participants, must rely on the observed outcomes for non-participants (the outcomes of the 1999 sample). A potential problem with this comparison is that the distribution of covariate (e.g. education) of the two samples has changed a lot between the two years. For instance, the proportion of single male households has increased, which might give a false impression of a reduction in the consumption, especially at the lower end of the distribution. Thus, a comparison of the realised outcomes of participants and non-participants is meaningless. Nevertheless, adjusting for the differences in their covariate distributions removes all selection bias and allows the estimation of treatment effects, provided all confounding characteristics are observed (Heckman 1990). It is, therefore, not assumed that each subsample's share of sample observations reflects the shares in the Danish population, but it is assumed that each subsample within the two samples is a random sample of its corresponding sub-populations, e.g. all uneducated single male Danish households in 2004 receiving 0-200,000 DKK and aged 25.

To adjust for the differences in their covariate distributions when comparing two consumption distributions, we use a methodology that enables comparison, even though the consumption distributions have different covariate distributions. The approach is similar to Machado and Mata (2005) for estimating unconditional densities based on any covariate distributions.

#### *A. Decomposing the changes in the distribution*

Denote by  $\hat{f}_t$  the sample estimator of the density of  $FaV$  (the log of consumption of fruit and vegetables) at time  $t$  based on the observed sample and by  $\hat{f}_t^*$  the model based estimator of the density based on the generated sample in step 4 below. The counterfactual density will be denoted by  $\hat{f}_t^{**}$ , for the density that would result in 2004 if all covariates had the distribution of 1999. The changes in the quantiles of the sample estimator,  $Q_t - Q_{t-1}$ , are decomposed by using a model-based estimator for the effect of the covariates and behaviour (coefficient).

in which the difference in the first parenthesis is the covariate effect and the difference in the second is the behavioural effect. Note that the behavioural/coefficient effect is the change in consumption when the covariate distribution is fixed, hence its name.

Besides the development in the population, we are also interested in the development in specific sub-populations and the relative change of two sub-populations. We can examine the change in the gap between two sub-populations during

the period, e.g. education group 3 and the uneducated, by subtracting the behavioural effect of one subgroup from the other:

By using the behavioural effect, the change in the gap during the period is adjusted for differences in socio-demographic composition in the two samples, as well as for the development in prices.

#### *The role of conditional distributions in multivariate simulation*

Write  $f(\mathbf{z})$  for the density of the random variable  $\mathbf{z}$ . When simulating from  $f(\mathbf{z})$  is the object of interest, but the distribution is unavailable, it might be possible to use the formula:

in which  $f(\mathbf{z}) = \prod_{j=1}^J f_j(\mathbf{z}_j | \mathbf{z}_{-j})$ . This is the case in this paper in which we do not know the distribution of the counterfactual density directly, but can obtain it indirectly through (6). The densities  $f_j(\mathbf{z}_j | \mathbf{z}_{-j})$  are available for simulation and simulation from  $f(\mathbf{z})$  is possible with the components in the product being generated sequentially using (6). In this paper, simulation from the conditional densities is obtained by quantile regression and the use of the inverse transformation method; the detailed process is described below.

#### *B. Unconditional densities implied by the conditional model*

Let  $FaV(t)$ , be consumption and  $\mathbf{z}(t)$  the covariates age, sex, income, household size and education at times,  $t=1999, 2004$ ,  $t=1999, 2004$ . Let  $g(\mathbf{z};t)$  be the sample density of the covariates at time  $t$ . Generating a random sample from the  $FaV$  density that would prevail in  $t$  if the model in (2) were true and the covariates were distributed as  $g(\mathbf{z};t)$  can be accomplished by:

1. Generating a random sample of size  $k$  from a uniform distribution:  $U(0,1)$ .
2. From the sample data at time  $t$  and each  $\{ \mathbf{z}_i \}$  estimate  $f(\mathbf{z}_i | \mathbf{z}_{-i})$  yielding estimates  $\hat{f}_i(\mathbf{z}_i | \mathbf{z}_{-i})$ ,  $i=1,..,k$ .
3. Generating a random sample of size  $k$  from  $\hat{f}_i(\mathbf{z}_i | \mathbf{z}_{-i})$ ,  $i=1,..,k$ .
4. Finally

A sample from  $f(\mathbf{z})$  and  $f(\mathbf{z})$  can be calculated by letting  $t$  equal 2004 and 1999, whereas a sample from the counterfactual density  $f(\mathbf{z}^*)$  can be accomplished by letting  $t=2004$  in step 1-2 and  $t=1999$  in step 3-4. To take the random nature of the procedure into account, the procedure is repeated many times and the median value of an interesting sample statistic, e.g. the first decile, is reported.

#### *C. Generating draws from the covariate distribution*

One question remains regarding how to calculate the distribution of the covariates,  $g(\mathbf{z};t)$ . Machado and Mata (2005)

suggest dividing continuous variables into groups, or bins<sup>37</sup>. After doing this to all the continuous variables they are left with a set of discrete variables and is then simple estimated by the frequency estimator, i.e. the maximum likelihood estimator. It is then easy to generate a sample from the estimated distribution of . In this paper, two deviations from their approach are taken. First, the continuous variable, price index, is not divided into bins. Instead, a draw from the covariate distribution without the price index is made, and a draw from the price index distribution is accomplished by treating price index as the response variable in a quantile model and following the procedure described above<sup>38</sup>.

Second, when generating draws from the discrete variables part of the covariate distribution, the frequency estimator is not used. Instead, an estimator which smooths over the discrete variables is used. Empirical studies have shown that estimators that smooth over discrete variables often give better predictions, especially when the number of cells is large relative to the number of observations. In this case it is also difficult to justify the frequency estimator on theoretical grounds (Li and Racine 2007, P. 120-121). This paper uses the estimator suggested in paper 3, which is a data driven method for selection of the smoothing parameters and is asymptotically superior to the frequency estimator and a recent suggested estimator for smoothing of multivariate discrete distributions (Ouyang *et al.* , 2006)<sup>39</sup>.

### 3. Empirical Analysis

#### 3.1 Mean regression

We conduct a series of consistent model specification tests for a mean regression of a linear parametric model with log gram as the response variable. The linear model starts with the simplest form that includes all of the six explanatory variables described in section 2.1. Two statistical software packages are used: NP for R developed by Hayfield and Racine (2008) and Quantreg for R developed by Koenker (2010). The test of the linear model gives a test statistic equal to 9.95003 and a corresponding P value, which is smaller than 2.22e-16. Thus, the null hypothesis that the simple linear model, which includes the six explanatory variables, is the correct model is rejected at any reasonable significance level. This is not unexpected as non-linearities could be for household size, for example, because of economies of scale in consumption<sup>40</sup>.

One might obtain a better fit of the data to the model if the first degree polynomials of the ordered/continuous variables are exchanged with higher order polynomials. Adding a quadratic term in age improved the test statistic and an additional cubic term even more so, but the p-value was still small. Adding a quadratic term for household size to the linear model also improved the test statistic relative to the linear model, but not nearly enough. The same was true for

---

<sup>37</sup> A continuous variable is split into ten bins and the length of the bins is decided by the 10<sup>th</sup> quantile, 20<sup>th</sup> quantile, ..., 90<sup>th</sup> quantile of the continuous variable, i.e. if an observation of the continuous variable is larger than the 10<sup>th</sup> quantile, but smaller than the 20<sup>th</sup> quantile, it is put into the second bin.

<sup>38</sup> After this draw from the covariate distribution it is finally possible to make a draw from the unconditional *FaV* distribution by doing step 1, 2 and 4 in the procedure described above.

<sup>39</sup> It is superior in the sense that the estimator is consistent just as these estimators and has the same variance when none of the discrete variables are uniformly distributed, i.e. they are asymptotically identical in this case, and has smaller variance when a subset of the discrete variables is/are uniformly distributed.

<sup>40</sup> It is well known that part of the food items purchased by households are discarded and it might be easier to minimize this quantity when the household size is larger.

the last continuous variable, namely the price index. This suggests that the best result would be obtained by including higher order polynomials for all three continuous variables. This gives the non-linear Model in table 3.1; the null hypothesis that this non-linear model has the correct specification could not be rejected at a high P value ( $P=0.49373$ ). This means that this non-linear model sufficiently models the non-linearities such as economies of scale in consumption for larger household sizes.

**Table 3.1** Consistent specification test of mean regression models.

Model		Test statistic	P value
Linear	Income, sex, education , household size, price index and age.	9.95003	$< 2.22e-16$ ***
	Terms added to Linear Model.		
Non-Linear	Age squared, the cubic of age , household size squared and price index squared.	-1.103015	0.49373

### 3.2 Quantile Regression

When referring to the Linear Model and Non-Linear Model in the following, the parametric structure is the same as in the mean regression section, except that the dependent variable is a quantile, instead of the mean. It is practically impossible to test the Non-linear Model and the Linear Model over all quantiles. Instead a range of quantiles were selected for the testing of the parametric Non-Linear quantile regression model and the parametric Linear quantile regression Model (see appendix). The tests revealed that the null hypothesis of correct specification could not be rejected for the parametric Non-Linear models and that this was not the case for the Linear Models (see Table 3.2 for the median regression test and the appendix for the rest of the tests).

**Table 3.2** Consistent specification test of median quantile regression models.

Model	Test statistic	P value
Linear	5.92391132717395	0
Non-Linear	-1.181218993	0.5714285714

#### *Counterfactual decomposition*

This section analyses the development in the Danish population of FAV purchases during the period 1999-2004 using the counterfactual decomposition (CFD) methodology.<sup>41</sup> As presented in table 3, the 10<sup>th</sup> percentile change in the data sample from 1999 to 2004 was negative (-0,145), although the actual change was likely to be near zero, as the behavioural effect is close to zero, i.e. the 10<sup>th</sup> and 90<sup>th</sup> percentile estimates have opposing signs and the median is close to zero.

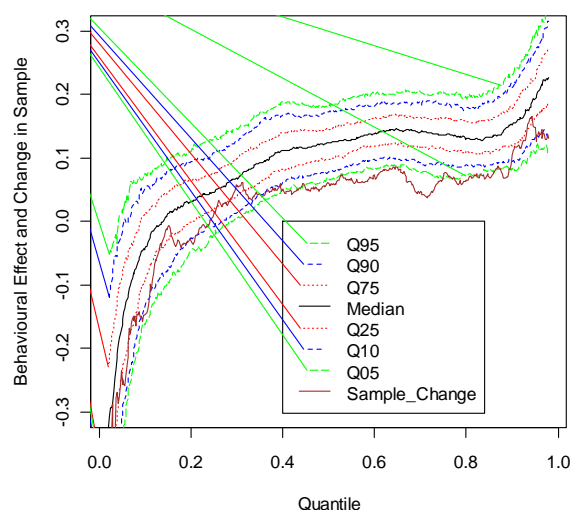
<sup>41</sup> The objective of the “6-a-day” campaign is to increase the consumption of FAV. Thus, success can be measured in the quantity consumed. The analysis is therefore based on the quantity purchased. Because of its apparent superiority, the Non-Linear Model is used in the counterfactual decomposition.

**Table 4** CFD of purchases of log gram of FAV using Non-Linear Model and equation 4.

	10 <sup>th</sup> quant.	25 <sup>th</sup> quant	Median	75 <sup>th</sup> quant.	90 <sup>th</sup> quant.
Sample 04	10,963	11,532	12,029	12,463	12,776
Sample 99	11,108	11,517	11,968	12,387	12,669
Change <sup>a</sup>	-0,145	0,015	0,060	0,076	0,106
Behavioural <sup>b</sup>	-0.042 (-.136, .052)	0,046 (-.007, .107)	0,123 (.078, .173)	0,136 (.093, .185)	0,151 (.095, .211)
Covariate <sup>b</sup>	-.109 (-.199,-.001)	-0.073 (-.137,-.013)	-.062 (-.114,-.019)	-0.050 (-.093,-.003)	-.066 (-.126, -.020)

<sup>a</sup> Change is the development in the data sample from 1999 to 2004. <sup>b</sup> Median (10<sup>th</sup> and 90<sup>th</sup> quantiles in parenthesis) of behavioural and covariate effects.

Table 4 suggests that the actual change, or behavioural change, is about zero at lower quantiles and positive at higher quantiles. The entire distribution is drawn in figure 1 by way of confirmation. It can be seen that, apart from the smallest quantiles, more than ten percent of the simulations have positive estimates of the coefficient effect<sup>42</sup>, thereby confirming that the low consuming household has not changed their behaviour. Looking at households with a higher consumption, it is clear that they have increased their intake, as more than 95 % of the simulations give positive behavioural effects for higher quantiles.



**Figure 1** Development in FAV purchases from 1999 to 2004 using equation 4 and sample change.

It is clear from figure 1 that the FAV purchases have increased, the proportional growth being larger at the higher quantiles.

Note that the median of the simulations of the behavioural effect is larger than the sample change at all quantiles, suggesting that the sample change underestimates the actual population change. This result is similar to Capacci *et al.*

<sup>42</sup> The 90<sup>th</sup> percentile of the simulations, i.e. Q90, is positive apart from the smallest quantiles.



(2010) who provide an *ex-post* assessment of the UK 5-a-day information campaign and find that all impacts are larger than those observed by simply comparing pre-policy and post-policy intakes.

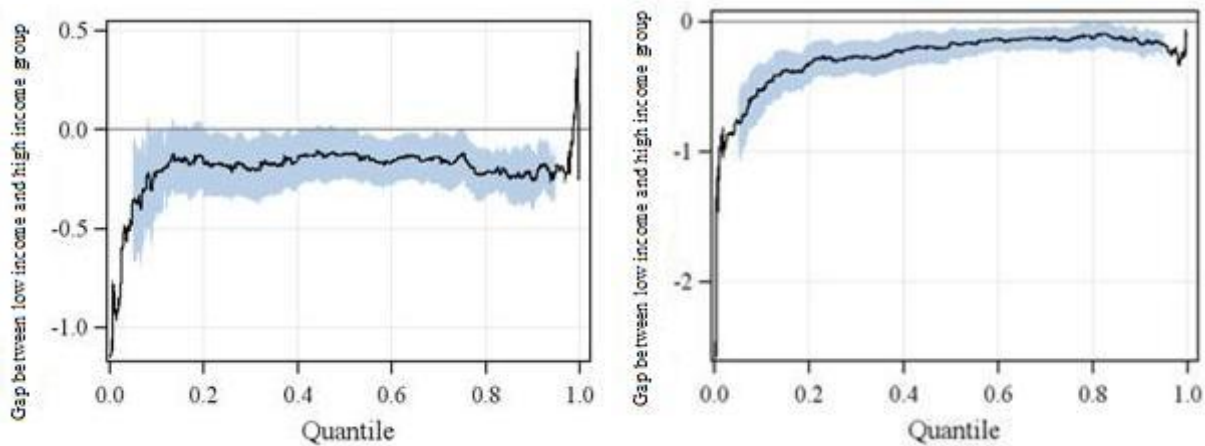
The status quo result of the low consuming groups, coupled with increased and proportional higher consumption at higher quantiles, implies increased inequality in the consumption of FAV.

A previous published study, based on diaries containing records of stated consumed FAV (Fagt *et. al.*,2008), also supports this hypothesis,.

We have now established that inequality in FAV consumption increased during the period 1999-2004, and are now ready to explore the determinants of this development.

#### *Development in socio-demographic differences*

The quantile process in figures 2 visualises the estimated gap in FAV purchases between low and high income groups at different quantiles, i.e. the parameter estimates for income group 1 of the quantile regression by quantile and its 95% significance bounds<sup>43</sup>. The plotting of the consumption of FAV for the low income group (income1) relative to the high income group (income3), shows that households with a low income purchase a smaller quantity than the high income group. In 2004 (1999), as illustrated, the consumption gap between income groups increases when moving from the 0.4(0.1) quantile and down through the distribution. This effect implies that the distribution in consumption of the low income group is more dispersed than it is for the high income group<sup>44</sup>.



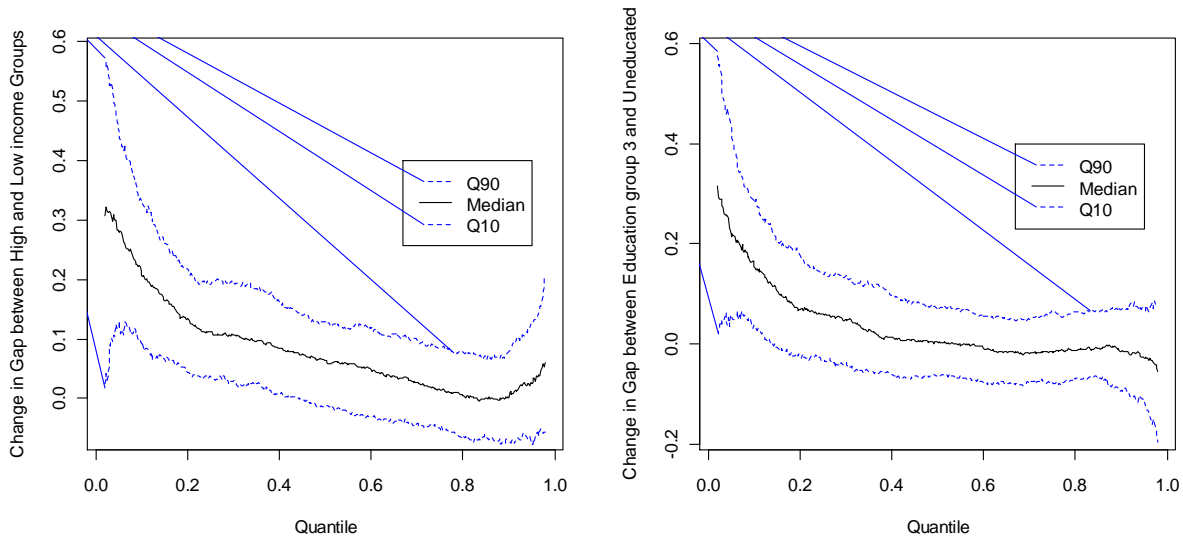
**Figure 2** Quantile process and 95 % confidence bounds. Estimated gap in FAV purchases in grams between low income and high income groups, i.e. the natural logarithm of low income FAV purchase minus the natural logarithm of high income FAV purchase by quantile. 1999 (left) and 2004 (right).

Households in the lower end of this distribution are hit twice. As they belong to the low income group they will spend less, and, since their distribution is much more dispersed, households in the lower end of the distribution spend much less.

<sup>43</sup> When the estimates are close to zero it can be interpreted as the percentage deviance, e.g. in figure 2  
<sup>43</sup>, where FAV is the purchases of fruit and vegetables measured in grams. This is because of the approximation  $\ln(1+x) \approx x$ , when  $x$  is small.

<sup>44</sup> A lower proportion of households, which belong to the low income group will contribute towards reduced inequality.

The comparison of the development in the income gap in Figure 2 does not take into account the change in the distribution of the socio-demographic variables and the development in FAV prices. We therefore plot the development over time in the FAV gap across income by utilising equation 5.



**Figure 3** Development in FAV purchases from 1999 to 2004 between income group 3 and income group 1 (left) and between Education group 3 and Uneducated (right).

It is clear from the left part of Figure 3 that the low income group has fallen behind the high income group across the entire distribution during the period, but especially so among households with low consumption. This means that the inequality within the low income group has increased more than it has within the high income group from 1999 to 2004. We have seen from figure 2 that low income groups are behind the high income group and figure 3 reveals that this tendency has been exacerbated.

This development is a cause for concern. The marginal health return to consuming more FAV is larger in the lower end of the distribution and the consumption might be increased if the household was given money, moving it to the high income group. A pressing question is whether this group has become relatively more income constrained during 1999-2004? An interpretation of this development is that households disliking FAV - belonging to the lower quantiles- seem to sacrifice FAV spending relatively more when their income is more constrained. The low income group could have become more income constrained because of increased prices and new goods and services which compete for the household's money<sup>45</sup>.

The right part of Figure 3 reveals that the education gap in purchases of FAV has increased among the low consuming groups. This supports the hypothesis that the information campaign has not worked on the uneducated low consuming segment of the population. The reason for the failure to increase the education gap among the high consuming segment of the population might be that this segment already knew the benefits of FAV consumption, which is why they were in the high consuming part to start with, and both education groups with high consumption intensity therefore have a similar response to the campaign.

<sup>45</sup> The low income group is defined by the same income bracket 0-200.000 DKK in both years.

The development in the income and education gap seen in figure 3 and the population change in figure 1 have been adjusted for prices, which means that this change is not caused by changes in prices. One is therefore inclined to believe that other factors such as an increase in information flow of the health benefits of FAV consumption and/or the 6-a-day campaign have driven the development.

#### *Impact of price development*

Figure 4 shows the difference between calculating the behavioural effect in the CFD with and without using a price index. Adjusting for price means that the prices are kept at the 1999 level. The difference, therefore, measures how much larger the change in the demand would have been from 1999 to 2004 if the prices in 2004 had been kept at the 1999 level, i.e. price responsiveness/sensitivity. The negative relationship in figure 4 reveals that low consuming groups are more price sensitive<sup>46</sup> and that high consuming groups would not have increased the gap to the low consuming groups to as large an extent if the price level in 2004 had been at the 1999 level. Thus, the price development has contributed to increased inequality.

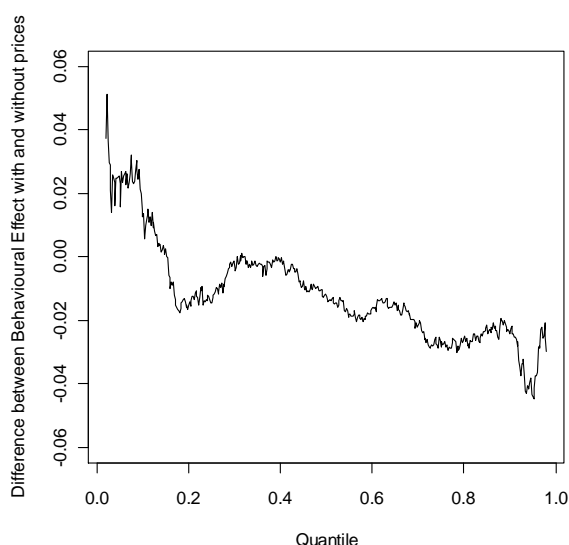


Figure 4 Measure of how much larger the change in the demand would have been from 1999 to 2004 if prices in 2004 had been kept at the 1999 level by quantile.

The measure in Figure 4 is similar in spirit as the difference in difference estimator, as used in e.g. Schreyögg *et al.* (2010), as they both subtract the difference which is not related to the price change/treatment.

<sup>46</sup> Note the possible problem of endogeneity, i.e. prices might not be exogenous because the price used in the analysis is determined by the goods purchased by the individual household and each household's price is therefore, to some extent, determined by its purchase behaviour. However, the quantile regression results of Ronning *et al.* (2004), which were based on a similar type of data (German GfK consumer panel), found that the shape of price elasticities "shows very similar price responses" when using an instrument variable approach. This then addresses the endogeneity problem, compared to what would be the case if an instrument variable approach was not used, thereby justifying not using an instrument variable approach.

## 4. Discussion and conclusion

Stewart *et al.* (2003) find that low income households spend less on FAV than high income groups over all quantiles.

This study sharpens their result, as we not only find that the cumulative distribution function of high income households' expenditure stochastically dominates low incomes,' we also find that this is the case for quantity.

There is data that suggest that FAV prices have the ability to modify consumption behaviour: Beydoun *et al.* (2008) found that lower FAV prices were positively associated with improved dietary quality and were protective against obesity, particularly among the near poor.

Price interventions in community-based settings, such as work sites and schools, have also shown promising results (e.g. French, 2003).

However, reducing prices as part of a policy to influence consumption behaviour can give unexpected results, because of the associated income effect and changes in relative prices of close substitutes. The effectiveness of a price policy can therefore depend on how it is implemented. For example, Hansen (2010, II) shows that a fat tax connected directly to the quantity of saturated fat contained in a food item is likely to be much more effective at decreasing purchases of saturated fat than a value-added tax (VAT). The results indicate that the VAT could have unfortunate policy implications for subpopulations because of unexpected price effects. The study therefore points to the importance of assessing whether the expected policy implications are aligned with measures of success. One such measure of success of a given policy is its ability to target the sub-population with the highest marginal health benefit. It is not obvious if a given policy has this trait, e.g. Bere *et al.* (2005) study the effect of a fee-based School Fruit programme and conclude that the School Fruit Programme appears to increase the intake among the subscribers, but also increases inequality by increasing the gap in FAV consumption among subscribers and non-subscribers.

One might fear that the response to a policy that decreases FAV prices would be that consumers with a high intensity of FAV purchases would respond by increasing consumption and households with a higher marginal health benefit with increased FAV consumption, i.e. low consuming groups would not increase consumption, which apparently has been the effect of the 6-a-day campaign.

This study, however, indicates that such fears might be unfounded, as low consuming groups had a higher price responsiveness, which could imply that the observed increase in the gap of FAV consumption between low and high consuming groups would have been smaller if prices had been kept at the lower 1999 level in 2004, hereby counteracting the inequality creating effect of the changes in behaviour, which could be attributed to the 6-a-day campaign.

The analysis provides evidence that supports the theory that information campaign do not work on the low consuming and uneducated. However, this conclusion may be slightly premature as it is worth remembering that the information campaign is focused on the educated and "ready to change" part of the population. The reason for this was an expectation about behaviour and a belief or theory that this group was more motivated to respond to dietary recommendations, which is the same as saying that the uneducated are not (as) responsive to dietary recommendations. This theory in turn shaped the information campaign, which focused on the educated, hence neglecting the uneducated in the process. Finally, an increase in the education gap among the low consuming groups was observed, and thus, the predicted behaviour was observed. Whether the low consuming, low education groups' non-responsiveness to the

campaign would have been observed if the information campaign had been more neutral, e.g. not directed towards any specific sub-groups, is an open question. Had this subgroup responded more favourably to a neutral information campaign, it could be concluded that the theory of non-responsiveness of the uneducated to information had become self-fulfilling (Ferraro *et al.*, 2005), i.e. by shaping the information campaign, thereby creating the behaviour it predicted. In conclusion, it is entirely possible that the theory of non-responsiveness of the uneducated to information and the actions based upon it can have contributed to increased inequality in health.

In this paper, we suggested smoothing over the discrete variables when using the method suggested in Machado and Mata (2005) when comparing distributions. We found that it was important to adjust the samples, because the sample change underestimated the actual population change.

Research limitations consist of possible inaccuracies in the measurement of FAV. Although data from the market research institute GfK Denmark provide an objective measure of purchasing behaviour, and this type of data have less response bias relative to self-reported dietary measures (Hebert *et al.* (2008), Kristal *et al.* (1998)), the purchase diaries represent household-level purchase data and not the food consumption of individuals within households. However, studies have shown significant positive correlations between household-level food purchases and dietary intake by individuals (Eyles *et al.* (2010) and Ransley *et al.* (2001) ).

This study reveals age, sex, income and educational differences in purchases of FAV. Inequality in education and income are spreading to the health area.

Relatively cheaper FAV would benefit low income groups more, because spending on FAV amounts to a much larger part of their available budget.

The study revealed that the behavioural change is larger than the sample change at all quantiles suggesting that the impact of the 6-a-day campaign is larger than what can be observed by simply comparing pre-policy and post-policy intakes. This result is similar to a recent study of the UK 5-a-day information campaign.

There was no change in the consumption among the lowest consuming groups, a finding that is supported by another recent study. This is an important result, as households in the lower end of the distribution have higher marginal health returns to consumption, and it becomes even more important from a government perspective if inequality is a concern. Arguably, some of this development could be attributed to low income groups becoming relatively more income constrained, since the gap to the high income group has grown considerably at the lower end of the distribution. This also means that the inequality has increased more within the low income group, than within the high income group.

The analysis revealed that the 6-a-day campaign could have contributed to increased inequality in health, which is in conflict with one of its policy goals and it has definitely not decreased the inequality in FAV consumption.

This study indicates that an information campaign directed toward the educated is unlikely to achieve the policy goals of the 6-a-day campaign, since uneducated low consuming groups have not responded well.

Changing the focus of the information campaign towards low consuming groups might be worthwhile, considering also

the higher health return of increased consumption of this group. However, such a policy which basically encourages individuals to consume more costly foods, might not be effective if standing alone, because this group is highly sensitive to price. A successful policy for public health should therefore emphasise the economics of food choice.

## 5. References:

- Andersen LT, Jensen H, Haraldsdóttir J. Typiske vægte for madvarer. (Typical weights for food items ). Scand J Nutr 1996; 4(Suppl. 32):S129-S152.
- Bere E., Veierød M. B., Klepp K.(2005) The Norwegian School Fruit Programme: evaluating paid vs. no-cost subscriptions. Preventive Medicine. Volume 41, Issue 2, August 2005, Pages 463-470.
- Beydoun M. A., Powell L. M. and Wang Y. (2008) The association of fast food, fruit and vegetable prices with dietary intakes among US adults: Is there modification by family income? Social Science & Medicine Volume 66, Issue 11, June 2008, Pages 2218-2229.
- Bremner, P, Dalziel D., and Evans, L. Evaluation of the 5 a day programme. Final report. Evaluation & Research Team. Big Lottery Fund: London; 2006.
- Capacci, S., Mazzocchi, M., Five-a-day; a price to pay: an evaluation of the UK program impact accounting for market forces, *Journal of Health Economics* (2010), doi:10.1016/j.jhealeco.2010.10.006.
- Danish Cancer Society (2006) KRÆFTENS BEKÆMPELSE. Omnibusundersøgelse om frugt og grønt. Datarapport. Jobnr. 3763. August 2006. Søren Pedersen, Dorthe Ibinger. TRANBERG MARKETING REKOMMANDATION Accessed on 17.11.10: <http://6omdagen.dk/dokumentation/tranberg/Aug2006rap.pdf>.
- European Commission (2005) Green Paper of 8 December 2005. "Promoting healthy diets and physical activity: a European dimension for the prevention of overweight, obesity and chronic diseases" (COM (2005) 637 final).
- European Commission (2007) WHITE PAPER ON - A Strategy for Europe on Nutrition, Overweight and Obesity related health issues. Brussels, 30.5.2007. COM (2007) 279 final.
- European Commission (2008) Working Paper on Public-Private Partnerships for Health, NOVEMBER 2008. Accessed on 16.10.10: [http://ec.europa.eu/health/ph\\_determinants/life\\_style/nutrition/documents/ev20081028\\_wp\\_en.pdf](http://ec.europa.eu/health/ph_determinants/life_style/nutrition/documents/ev20081028_wp_en.pdf)
- Eurostat (2010) The European Health Interview Survey (EHIS). Eurostat, Health & Safety Statistics. January 2010. Accessed on 18.10.10: [www.unece.org/stats/documents/ece/ces/ge.13/2010/wp.9.e.ppt](http://www.unece.org/stats/documents/ece/ces/ge.13/2010/wp.9.e.ppt)
- EHES (2009) EHES - Measuring the Health of Europeans. [http://www.ehes.info/flyied\\_04022010.pdf](http://www.ehes.info/flyied_04022010.pdf). Accessed on 17.11.10.
- Eyles HC, Jiang Y, Mhurchu C. N.(2010) Use of household supermarket sales data to estimate nutrient intakes: a comparison with repeat 24-hour dietary recalls. J Am Diet Assoc 2010;110:106–10.
- Fagt S., Jensen A. B., Matthiessen J., Groth M. V., Christensen T., Trolle E.(2008). Dietary habits of Denmark 1995-2006. Status and development with focus on fruits, vegetables and added sugar. ISBN:

- Ferraro F., Pfeffer J., Sutton R. I. (2005) Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management Review*, 2005, Vol. 30, No. 1, 8–24.
- French S. A. (2003) Pricing Effects on Food Choices. Symposium: Sugar and Fat-From Genes to Culture. *J. Nutr.* 133:841S-843S, March 2003.
- Ministry of Food, Agriculture and Fisheries/Fødevaredirektoratet (2001) *6 om dagen – Overordnet kampagnestrategi*. Ref.: LSL/ETR. J.nr.: KOMS 3520-1/00. 10. Januar 2001. Fødevaredirektoratet
- Groth and Fagt (2003). Danskernes kostvaner. Måltidsvaner, holdninger, sociale forskelle og sammenhæng med anden livsstil. Fødevaredirektoratet, 2003.
- Gordon R, McDermott L, Stead M, and Angus K. (2006) The effectiveness of social marketing interventions for health improvement: What's the evidence? *Public Health* 2006;120; 1133-1139.
- Havas S. et al. (1995) 5 A Day for Better Health-Nine Community Research Projects To Increase Fruit and Vegetable Consumption. *Public Health Reports*. January-February 1995, Vol. 110, No. 1 71. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1382077/pdf/pubhealthrep00056-0070.pdf>.
- Heckman J. (1990). Varieties of selection bias. *American economic review. Papers and proceedings*, 80, 313-318.
- Gustavsen G. W. and Rickertsen K. A censored quantile regression analysis of vegetable demand: The effects of changes in prices and total expenditure. *Canadian journal of agricultural economics* 54, 631-645.
- Hayfield T. and Racine J. S. (2008) Nonparametric Econometrics: The np Package. *Journal of Statistical Software*. July 2008, Volume 27, Issue 5.
- Haraldsdóttir J, Holm L, Kristiansen M et al. *Befolkningens fødevarerforbrug I 2004 . Food Consumption in Denmark 2004 compared to results from 1995, 1998 and 2001. ISSN. Nr. 0909-9859.*
- Hebert J. R., Hurley T. G., Peterson K. E., Resnicow K., Thompson F. E., Yaroch A. L., Ehlers M., Midthune D., Williams G. C., Greene G. W., and Nebeling L. (2008) Social Desirability Trait Influences on Self-Reported Dietary Measures among Diverse Participants in a Multicenter Multiple Risk Factor Trial. *J. Nutr.* 138: 226S–234S, 2008.
- Hsiao, C., Li, Q., & Racine, J. (2007). *Consistent Model Specification Tests with Mixed Categorical and Continuous Data*. *Journal of Econometrics*, 140, 802-826.
- José A. F. Machado and José Mata. *Counterfactual Decomposition of changes in wage distributions using quantile regression*. *J. Appl. Econ.* 20: 445-465 (2005)
- Koenker, R.W. and G.W. Bassett (1978), “Regression quantiles”, *Econometrica*, 46, 33-50.
- Koenker R. (2010) Quantile Regression. Package ‘quantreg’. September 29, 2010  
Accessed on 17.11.10: <http://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
- Kristal A. R., Andrilla C. H. A., Koepsell T. D., Diehr P. D (1998) Dietary assessment instruments are susceptible to intervention-associated response set bias. *J Am Diet Assoc* 1998;98:40–3.
- Li, Q. and J.S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- National Accounts 2005
- Published by Statistics Denmark, May 2007. Pdf-edition: ISBN 978-87-501-1609-7. ISSN 1901-6239
- National Accounts 2007

Published by Statistics Denmark, February 2009. Pdf-edition: ISBN 978-87-501-1752-0. ISSN 1903-4563.

Ouyang, D. and Q. Li and J.S. Racine (2006), "Cross-validation and the estimation of probability distributions with categorical data," *Journal of Nonparametric Statistics*, 18, 69-100.

Pomerleau J, Lock K, Knai C, and Mckee M. Interventions designed to increase adult fruit and vegetable intake can be effective: A systematic review of the literature. *Journal of Nutrition* 2005;135; 2486-2495.

Racine, J.S. (2006), "Consistent specification testing of heteroskedastic parametric regression quantile models with mixed data," manuscript.

Ransley J. K., Donnelly J. K., Khara T. N., Botham H., Arnot H., Greenwood D. C. and Cade J. E. (2001). The use of supermarket till receipts to determine the fat and energy intakes in a UK population. *Public Health Nutr* 2001;4:1279–86.

Government /Regeringen (2002) *Sund hele livet - de nationale mål og strategier for folkesundheden 2002-10*.

Regeringen September 2002. Udgivet: September 2002. Udgivet af Indenrigs- og Sundhedsministeriet, Slotsholmsgade 10-12, 1216 København K. ISBN 87-90227-67-0, trykt version, ISBN 87-90227-69-7, elektronisk version.

Schreyögg J. and Grabka, M. M. (2010). Copayments for ambulatory care in Germany: a natural experiment using a difference-in-difference approach. *The European Journal of Health Economics*. Volume 11, Number 3, 331-341, DOI: 10.1007/s10198-009-0179-9

Stewart H., Blisard N and Jolliffe D(2003). Do income constraints inhibit spending on fruit and vegetables among low-income households? *Journal of agricultural and resource economics* 28(3):465-480.

U.S Department of Health and Human Services (2004). *Conquering Obesity: The U.S Approach to Combating this National Health Crisis*. September 15, 2004.

Accessed on 17.11.10: <http://www.hhs.gov/asl/testify/t040915.html>.

WHO/FAO(2003) *Expert Report on Diet, Nutrition and the Prevention of Chronic Diseases*, Technical Report Series 916, 2003,WHO/FAO.

**WHO (2004) GLOBAL HEALTH RISKS. Mortality and burden of disease attributable to selected major risks.**

[http://www.who.int/healthinfo/global\\_burden\\_disease/GlobalHealthRisks\\_report\\_full.pdf](http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf)

Zheng, J. (1998), "A consistent nonparametric test of parametric regression models under conditional quantile restrictions", *Econometric Theory*, 14, 123-138.



## 5. Appendix:

### Kernel Consistent Quantile Regression Model Specification Test of non-linear model in year 1999.

Quantile $q$	statistic $I_n$	statistic $J_n$	P-value of the statistic $J_n$
0.05	-3.15241E-8	-0.96513852	0.5438596491
0.1	-9.029346E-8	-1.453838574	0.7268170426
0.15	-2.50126E-8	-0.286810348	0.1679197995
0.2	-4.248556E-8	-0.371876241	0.1679197995
0.25	-1.561757E-8	-0.118605641	0.1503759398
0.3	-6.660531E-8	-0.451524325	0.2255639098
0.35	-1.22252E-7	-0.75342487	0.343358396
0.4	-1.789905E-7	-1.044724182	0.5012531328
0.45	-2.708963E-7	-1.510647182	0.7468671679
0.5	-2.148033E-7	-1.181218993	0.5714285714
0.55	-2.849929E-7	-1.574324754	0.7819548872
0.6	-2.304973E-7	-1.310790512	0.6390977444
0.65	-1.767432E-7	-1.069564891	0.4987468672
0.7	-7.053248E-8	-0.463956051	0.2130325815
0.75	1.0830648E-8	0.0831714895	0.0927318296
0.8	1.8928473E-8	0.1732356573	0.0902255639
0.85	-4.10553E-8	-0.454768036	0.2080200501
0.9	-2.070483E-8	-0.320238206	0.1729323308
0.95	-2.669435E-8	-0.828188171	0.4035087719
<p>The test is produced using the function <code>npqcmstest</code> in the R software.</p> <p><code>npqcmstest</code> implements a consistent test for correct specification of parametric quantile regression models (linear or nonlinear) as described in Racine (2006) which extends the work of Zheng (1998). For specific details see the NP package in the R software.</p>			

# Paper 6: Non-linearities in consumers' demand response to price changes - the case of milk demand

---

Aslak Hedemann Hansen & Jørgen Dejgaard Jensen

## **Abstract**

This paper examines a hypothesis of an existence of non-linearities in consumers' demand response to price changes using Bayesian estimation techniques. The analysis extends the empirical demand analysis by searching for potential non-linearities in the response to price changes. In general, the empirical analysis suggests that the demand for drinking milk responds significantly to price changes. But, the analysis also shows some asymmetry in this response to price changes. In particular, price decreases tend to trigger larger demand responses than price increases.

## **Keywords:**

Non-linearities in demand response, Monte Carlo Markov Chain, milk demand.

## 1. Introduction

Obesity is recognised as an increasing problem in most parts of the world (WHO, 2006) and research being carried out in many different fields in order to develop measures to prevent further growth in obesity.

One of the possible instruments that have been put forward in several studies is the use of food taxes, for instance, by subsidising nutritious foods and increasing taxes on unhealthy foods (Nordström & Thunström, 2009; Gustavsen & Rickertsen, 2006 ; Jensen & Smed, 2007)

Many of the empirically based studies in this field build on price elasticities that have been estimated econometrically on the basis of price and demand fluctuations in historical data, assuming a linear relationship between percentage price changes and percentage demand responses.

However, the realism in using such a log-linear approach to tax-induced price changes has hardly been investigated when it comes to food demand relationships. As food consumption is, to a large extent, driven by socio-cultural norms (Renzaho, 2004), habits (Naik & Moore, 1996), practical and convenience considerations (Warde, 1999), there may be reason to expect some rigidity in the demand response to price changes. For example, it may be expected that minor price changes will have practically no effect on consumption, whereas price changes beyond a certain threshold, may induce an economic incentive to change behaviour that exceeds the perceived transaction costs associated with changing the food consumption.

Several studies have shown positive health effects of consuming milk and other dairy products. For instance, dietary patterns characterised by increased dairy consumption have a strong inverse association with Insulin Resistance Syndrome (IRS) among overweight adults and may reduce the risk of type 2 diabetes and cardiovascular disease (Pereira *et al.*, 2002).

Also, children who avoid milk are more likely to have poor bone health, are prone to bone fracture and are associated with small stature (Goulding *et al.*, 2004; Black *et al.*, 2001). The scientific evidence which links dairy foods and a reduced risk of osteoporosis have resulted in the (2005) Dietary Guidelines for Americans, which recommend 3 servings of milk products per day to reduce the risk of low bone mass and to contribute important nutrients that may have additional health attributes beyond bone health (Huth *et al.*, 2006).

Furthermore, milk constitutes a basic element of dietary behaviour for a significant share of the Danes (Fagt *et al.*, 2008).

The purpose of this study is to empirically investigate whether such rigidities and thresholds exist, the size and symmetry properties of such effects and the potential differences between different commodity categories.

This introduction is followed by a description of the methodology and data, whilst the subsequent section provides some key results of the analysis. The results are discussed in section 4, and the final section draws some conclusions and perspectives.

## 2. Methodology and data

### 2.1 Theoretical framework

The study takes its departure in standard microeconomic utility maximisation theory in which the consumer is assumed to maximise her utility  $U(x)$  subject to the budget constraint  $p'x = y$ , where  $x$  is a vector of consumed quantities of goods,  $p$  is a corresponding vector of prices, and  $y$  is the total budget available for consumption. Assuming

normal regularity properties of the utility function (monotonicity, concavity etc.), consumers' demand for different commodities responds to price changes, depending on the shape of the utility function. This can be represented by the own-price elasticity. If the demand curve is downward-sloping and concave, the price elasticity associated with large price increases will be larger than for small price increases, and the price elasticity associated with a large price decrease will be smaller than for minor price decreases. The latter could be interpreted as a saturation effect, i.e. even though prices fall considerably, there is a limit to how much the consumers desire to consume (van Heerde *et al.*, 2001, Fox *et al.*, 2004, Blattberg *et al.*, 1995). On the other hand, a convex demand curve can be interpreted such that the good is a necessity, in the sense that there is a lower limit to how much the consumer will consume, even for high price changes.

This theoretical model may be expanded by assuming that consumers have imperfect knowledge about own preferences or prices, but this knowledge depends on the amount of information,  $I$ , the consumer has obtained and processed (for example, the consumer does not know whether she likes a particular food before she has experienced it). As the gathering and processing of information implies a cost  $c(I)$ , for instance, time spent for monitoring prices, for acquiring necessary background knowledge to process available information, or for obtaining experience with different goods, some consumers base their current price estimation on heuristics (Gutenberg, 1976, Hruschka 2000, Gilbride and Allenby 2004, Kahneman, 1991) and consequently, they underestimate price changes (e.g. Blair and Landon 1981, Mobley *et al.*, 1988, Gupta & Cooper, 1992).

One implication of this expansion of the model is that, for small price changes, the consumer may not find it worthwhile to switch to other goods, because the costs of obtaining relevant knowledge for this decision may exceed the potential efficiency gain. This imperfect knowledge aspect may suggest a threshold effect, which implies that small price changes will trigger smaller demand responses (or no response) than larger price changes. Another interpretation of such an effect could be that the consumers exhibit some degree of brand loyalty (Blattberg *et al.*, 1995).

Together, the curvature of the utility function and the costs of information suggest that demand responses to price changes may follow a non-linear pattern, with potential asymmetry between effects of positive and negative price changes and the existence of a threshold for demand responses to occur. The exact shape of the demand response pattern is an empirical question, which is presumed to depend on the type of good considered. Wisniewski and Blattberg (1983) find that the demand curve fits an S-shaped function, which is consistent with both threshold and saturation effects, whereas Blattberg and Wisniewski (1987) find the demand curve to be convex, although based on a relatively limited range of price discounts. Fok *et al.* (1997) investigate demand responses for four soft drink brands and find a steeper demand curve for price decreases than for price increases, and that demand is relatively more sensitive to small price changes than to larger price changes.

Hypotheses:

- Are friction patterns symmetric with regard to positive versus negative price changes?
- Are friction patterns similar across different types of milk?

## 2.2 Empirical model

The above theoretical considerations suggest that demand responsiveness to price changes may belong to one of three regimes: “large” price decreases, “small” price changes, or “large” price increases. These regimes are bounded by two thresholds:  $\delta_L < 0$ , and  $\delta_U > 0$ . That is, observations where,  $\Delta P_{it} < \delta_L$  fall in the “large price decrease” regime, observations where,  $\Delta P_{it} > \delta_U$  belong to the “large price increase” regime, and observations with,  $\delta_L \leq \Delta P_{it} \leq \delta_U$  are categorised as “small price change” observations. An implication of the above considerations is that the demand response to a price change may depend on which of the three regimes the observation belongs to.

Two general approaches have been proposed in the literature to estimate such threshold models: a latent variable structural-form approach and a reduced-form approach. In the latent variable approach, as specified in the friction model by Rosett (1959), observed behaviour (e.g. a change in a consumer’s purchase of a commodity),  $\Delta x_{it}$ , reflects the value of an unobservable latent variable  $\Delta x_{it}^*$ , which in turn depends on a number of underlying variables, including price change,  $\Delta P_{it}$ . However, the relationship between the observed and the latent variable is censored in the sense that it depends on the regime of the latent variable.

$$\begin{aligned}
 \Delta x_{it}^* &= \beta' \Delta P_{it} + u_i \\
 \Delta x_{it} &= \Delta x_{it}^* - \delta_L \text{ if } \Delta x_{it}^* < \delta_L \\
 \Delta x_{it} &= 0 \text{ if } \delta_L \leq \Delta x_{it}^* \leq \delta_U \\
 \Delta x_{it} &= \Delta x_{it}^* - \delta_U \text{ if } \Delta x_{it}^* > \delta_U \\
 \delta_L < 0, \delta_U > 0
 \end{aligned} \tag{1}$$

Various refinements to the Rosett model have been proposed, including e.g. the model by Omoro & Miyawaki (2009), where thresholds are allowed to vary with covariates.

On the other hand, the reduced-form approach specifies the response function as a piecewise linear function of the dependent variable (Fok *et al.*, 2007), with a corresponding likelihood function, such as

Where

The term  $\ln \pi(\theta)$  is the sum of the log of the prior densities and the term  $\ln \pi(y)$  is the log

likelihood.

The empirical model used in this study builds on formulation (2), and includes lagged values of the consumed quantities (Pauwels *et al.*, 2006). In addition to the positive and negative threshold breakpoint, we can also consider potential asymmetry between positive and negative “small price changes,” i.e. a third change point  $\delta_0$  (close to zero). Hence, our empirical model includes four intervals for the price response parameter  $\beta$ , reflecting different combinations of positive and negative price changes, as well as “small” and “large” price changes. Hence, the empirical model consists of equations of the form outlined in expression (2.1) and (2.3)

(2.3)

In particular, single-equation models are specified for four types of drinking milk: skimmed milk (0.5% fat), semi-skimmed milk (1.5% fat), mini-milk (milk with a fat content like skimmed milk, but with a taste close to semi-skimmed milk), and whole milk (3.5% fat).

## 2.3 Data and estimation

For the study, we use shopping data from the Danish GfK ConsumerScan household panel (GfK, 2010). The dataset contains detailed weekly purchase recordings of a range of nondurable consumer goods, including a large number of food and beverage products, for 2000 Danish households. For most goods, the collected information includes: purchased quantity, expenditure, package size and type of packaging, brand and specific characteristics (e.g. flavour, low-fat, organic, etc.). For each household in the panel, background information is available, including e.g. household size, age and sex of household members, income, education, and a number life-style indicators, such as type of dwelling, readership of journals or magazines, possession of household equipment, etc.

The panel structure of the GfK dataset enables an analysis of key parameters in consumer behaviour, including consumer responses to changes in price conditions, at a fairly detailed level, and with a distinction between different types of households. Hence, the data set appears appropriate for analysing the theoretical model outlined above.

In this study, we focus on demand behaviour for four different types of milk: whole-milk, semi-skimmed milk, skimmed milk and taste-improved skimmed milk (so-called 'mini-milk'). In addition to being important elements in the Danes' diets, GfK data on milk consumption is also of good quality, with well-defined products and fairly consistent product definitions over time, and with a large number of observations. Descriptive statistics from the data are given in table 1.

Table 1. Descriptive statistics on milk consumption

	Skimmed milk	Mini milk	Semi-skimmed milk	Whole milk
N	35552	33720	67807	28244
Mean consumption (litres per week)	2,22	2,24	2,34	1,68
St.dev.	1,58	1,57	1,84	1,13
Mean price (DKK/litre)	5,96	5,77	5,83	6,82
St.dev.	0,63	0,70	0,73	0,74
1% Quantile price	3,95	3,35	3,65	3,95
99% Quantile price	7,05	7,50	7,47	8,31

Source: GfK Consumer Tracking panel

The non-linearities of the model outlined in expression (3), with combinations of non-linearities and thresholds, poses some challenges to the econometric estimation.

One data issue regarding estimation of the model is the fact that we only have household-level price recordings in weeks when purchases have actually taken place. Regarding the above model, this constitutes two problems. One problem is that households' decision not to buy a commodity may have been a result of prices, and hence omitting non-purchase observations may yield biased estimations of the relationships between price and consumption. Another problem is the dynamic structure of the model in (3), where information on price changes between consecutive weeks is needed.

We attempt to solve this problem by approximating household-level milk prices in a certain week with the average corresponding milk price for all consumers. If household-level milk prices are correlated across households, the week-to-week changes in the milk prices can be reasonably well be approximated by week-to-week changes in the corresponding average milk prices.

Having constructed average price indices for the four milk types, respectively, we estimate model (2.1) and (2.3) using a Bayesian simulation technique, in terms of a Monte Carlo Markov Chain method. According to this iterative approach, a sub-set of parameters is estimated conditional on the data and other predetermined sub-sets of parameters (either by random draws from a prior distribution or from estimation), and conditional on the estimation of this sub-set of parameters, another sub-set can be estimated, and so on. For this purpose, parameters were grouped into three sub-sets:  $(\beta_1, \beta_2)$ , and (s2). After a large number of iterations, it is possible to establish statistical distributions for the model parameters. The models for the four milk types are estimated using SAS PROC MCMC.

### 3 Results

#### 3.1 Results for all consumers

It was attempted to estimate the model outlined in expression (2.1) and (2.3) for four types of milk: skimmed milk, 'mini' milk, semi-skimmed milk and whole milk, with three change points for the  $\beta$  parameters. However, the applied MCMC estimation procedure was only able to identify and estimate one change point parameter in each equation.

Hence, econometric results for models with only one change point for all households are reported in table 2.

As an example, beta1 and beta2 represent the effect of 1 DKK/litre price change on the demand for milk (measured in liters), for price changes below or above the change point, respectively. Hence a price reduction of 1 DKK/litre (below the change point of -0.20 DKK/litre) leads to a demand increase of 0.79 litres, whereas a price increase of 1 DKK will reduce the weekly demand by 0.31 litres for an average household.

Table 2. Econometric estimation results

	Skimmed milk	Mini milk	Semi-skimmed milk	Whole milk
Alpha	0.040 (0.012)	-0.109 (0.009)		
Delta	-0.200 (0.017)	0.075 (0.011)	-0.135 (0.019)	-0.264 (0.046)
Gamma	-0.750 (0.004)	-0.751 (0.004)	-0.833 (0.003)	-0.825 (0.005)
Beta1	-0.7890 (0.0269)	-0.6180 (0.0171)	-0.4785 (0.0168)	-0.3507 (0.0188)
Beta2	-0.3136 (0.0254)	0.0065 (0.0184)	-0.1872 (0.0127)	-0.0795 (0.0127)
S2	3.149 (0.018)	2.902 (0.016)	3.799 (0.017)	1.598 (0.010)

Note: Standard deviations in parentheses

For all milk types, the results show a clear asymmetry in the short-term demand response to price changes, in that negative price changes (below the change point) trigger a demand response that is more than double the response of positive price changes. Hence, the estimation results suggest a kinked demand curve for the respective milk products, as illustrated in figure 1.

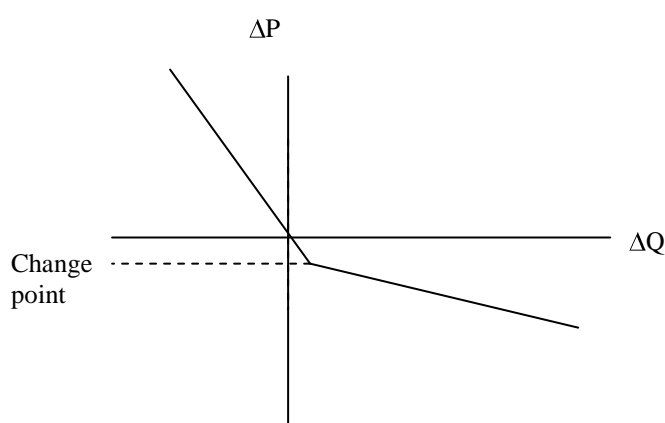


Figure 1. Illustration of estimated demand relations (skimmed, semi-skimmed or whole milk)

The delta coefficient represents the change point for the  $\beta$  parameters. These coefficient estimates are statistically significant (assuming that the parameter estimates are asymptotically normally distributed, which is supported by visual

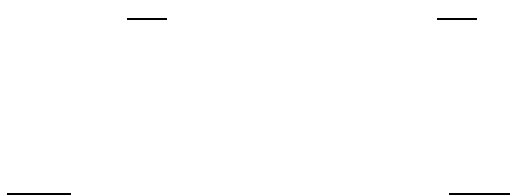


inspection of density plots from the estimation output). For skimmed milk, semi-skimmed milk and whole milk, the parameter is negative, which suggests that, for minor price decreases, the demand response will be symmetric with that of corresponding price increases, whereas the demand effect may be relatively stronger for larger price decreases.

For skimmed, semi-skimmed and whole milk, the results suggest that the demand response to price decreases depends on the size of the price decrease. Specifically, the price of skimmed milk decreases larger than 0.20 DKK/litre trigger a demand response that is relatively stronger than smaller price decreases. In contrast, for positive price changes, no such threshold was found with regard to these three milk types. On the other hand, no such effects could be found for mini milk, as far as price decreases are concerned.

It should be noted that the demand response parameters are estimated on the basis of week-to-week changes in price and demand, and hence represent very short-run responses. In particular, it could be imagined that special offers on milk in one week, with a return to the normal price in the following week, may have an important impact on the parameter estimates. And despite the relatively short durability of fluid milk, it might be considered likely that consumers would tend to utilise such special offers to fill their refrigerators with milk by the end of a special-offer week, and these purchases would last almost until the end of the subsequent week, whereas the reverse behaviour is not considered likely in the case of price increases.

The identified asymmetry in the demand response to price changes deserves a comment regarding the consistency with economic theory. This may be examined further by the following derivations:



Consider now a temporary price decrease (below change point) in one week ( $t-1$ ), followed by a restoration of the original price level in the following week ( $t$ ). The total effect of this price scenario on the demand in period  $t$  is given by:

$$\frac{\partial Q}{\partial P} \Delta P + \frac{\partial Q}{\partial P} \Delta P$$

The two oppositely directed price changes should be expected to more or less offset each other, so that the demand in period  $t$  should be similar to the demand in a situation without this price drop. This is also the case with the considered temporary price drop.

On the other hand, if we consider the opposite scenario – a temporary price increase larger than the estimated change point (in absolute value) – the total effect does not come close to zero. In this case, the outcome is substantially different from that of a stable price during the two periods, which in principle raises concern for the theoretical consistency of the estimated model. It is, however, presumed that temporary (one-week) significant price drops are more frequent than one-week significant price increases, and hence the estimated model performs fairly reasonably for most of the

observations in the dataset.

The introduction of lagged price changes into the model might be considered – with asymmetric demand responses to these lagged price changes. Another way to re-specify the model could be to introduce asymmetry in the parameter associated with the lagged quantity variable. These suggestions may serve as inspiration for future work in the analysis of asymmetries in demand responses.

### 3.2 Results for households with small children versus other household

As mentioned in the introduction, early childhood nutrition is suspected to play an important role for nutrition, obesity risk and related health problems (Goulding *et al.*, 2004; Black *et al.*, 2001) over the entire lifetime. For this reason, it is interesting to investigate whether the milk demand behaviour of households with small children differs from that of other households, and in particular if households with children are more price responsive than other households.

In order to address this question, the above model has been estimated for two sub-samples of the household panel dataset: a sub-sample of households with children aged 0-6 years (10712 observations), versus a sub-sample of households without children in this age span (78832 observations). For each of these two sub-samples, the above model has been estimated using the econometric procedure described above. Results from this estimation are presented in table 3.

Table 3. Selected econometric estimation results, households with or without small children (0-6 years)

	Skimmed milk	Mini milk	Semi-skimmed milk	Whole milk
<b>Children</b>				
Delta	1.794 (0,193)	0,306 (0,409)	-1.878 (0,086)	0,078 (0,195)
Beta1	-0.5042 (0.0498)	-0.7568 (0.0597)	0.1254 (0.0885)	-0.3049 (0.0344)
Beta2	-0.0495 (0.1007)	-0.1195 (0.0797)	-0.5969 (0.0372)	-0.0488 (0.0335)
<b>No children</b>				
Delta	-0,195 (0,015)	0,034 (0,057)	-0,116 (0,022)	-0,600 (0,094)
Beta1	-0.6471 (0.0250)	-0.5372 (0.0185)	-0.4547 (0.0172)	-0.4299 (0.0309)
Beta2	-0.0786 (0.0206)	0.0198 (0.0336)	-0.1114 (0.0162)	-0.0681 (0.0131)

Note: Standard deviations in parentheses

Overall, the results of this partitioned estimation support the above-mentioned asymmetry between positive and negative price changes, with larger responses for price decreases than for price increases. But, the partitioned results also suggest some further nuances in the demand behaviours. For example, households with children exhibit a different demand behaviour with regard to semi-skimmed milk than for other milk types, whereas this is not the case for other households. Another noteworthy feature is the rather high positive change point for skimmed milk demand in households with small children, which suggests that, in these households, a linear demand relationship prevails unless substantial price increases occur.

Overall, the partitioned results do not reveal clear and systematic differences between households with and without

small children, except for the case of semi-skimmed milk, where the demand behaviour for the former group of households tends to deviate from the general pattern.

## 4 Discussion

It should also be noted that the estimated equations only contain own-price effects, which is partly due to high correlation between different milk prices and to technical estimation problems with substantially increased numbers of parameters. As some of the milk types are expected to be relatively close substitutes, this could be considered as a limitation to our analysis.

## 5 Conclusion

This paper analyses price responsiveness in milk demand, distinguishing four different types of drinking milk differentiated according to fat percent and taste. The present analysis brings the empirical demand analysis further by searching for potential non-linearities in the response to price changes.

In general, the empirical analysis suggests that the demand for drinking milk responds significantly to price changes. But, the analysis also shows some asymmetry in this response to price changes. In particular, price decreases tend to trigger larger demand responses than price increases.

An interesting finding in the empirical results is that there exists a change point (which is most often significantly different from zero), around which this asymmetry occurs. Hence, for some types of milk, e.g. skimmed milk, small price reductions have moderate demand effects, whereas price reductions beyond a threshold of 0.2 DKK/litre trigger substantially higher demand increases.

To the extent that the intake of milk has a role to play in the prevention of overweight and related disease risks, lower prices for milk may be considered (e.g. through a consumption subsidy or VAT exemption). And the results of this study indicate that large price reductions may be relatively more effective for that purpose than smaller price reductions.

## References

- Anderson W. (2007) An Alternative Event Study Methodology for detecting Dividend Signals in the Context of Joint Dividend and Earnings Announcements, Paper prepared for the AFAANZ Conference, Conrad Jupiter's Casino, Boldbeach, the Gold Coast, July 1-3, 2007
- Black R. E., Williams S. M., Jones I. E. & Goulding A. (2002). Children who avoid drinking cow milk have low dietary calcium intakes and poor bone health1–3. *Am J Clin Nutr* **September 2002** vol. 76 no. 3 **675-680**.
- Blair, Edward A. and E. Laird Landon Jr. ( 1981), "The Effects of Reference Prices in Retail Advertising, " *Journal of Marketing*, 45 (Spring), 61-69.
- Blattberg, Robert C. and Kenneth J. Wisniewski (1987), "How Retail Price Promotions Work: Empirical Results," marketing working paper No. 42, Graduate School of Business, University of Chicago.
- Blattberg R.C., Briesch R. & Fox E.J. (2005) How promotions work, *Marketing Science*, vol. 14, pp. G122-G132
- Dagenais M.G. (1975) Application of a Threshold Regression Model to Household Purchases of Automobiles, *Review of Economics and Statistics*, vol. 57, pp. 275-85.

Fagt S., Jensen A. B., Matthiessen J., Groth M. V., Christensen T., Trolle E.(2008).

Dietary habits of Denmark 1995-2006. Status and development with focus on fruits, vegetables and added sugar. ISBN: 978-87-92158-19-2 .www.food.dtu.dk.

Fok D., Franses P.H., & Paap R. (2007) Seasonality and non-linear price effects in scanner-data-based market-response models, *Journal of Econometrics*, vol. 138, pp. 231-251.

Fox E.J., Montgomery A.L., Lodish L.M. (2004) Consumer shopping and spending across retail formats, *Journal of Business*, vol. 77, pp. S25-S60.

GOULDING A. *et al.* (2004). Children Who Avoid Drinking Cow's Milk Are at Increased Risk for Prepubertal Bone Fractures. *J Am Diet Assoc.* 2004;104:250-253.

Granger C.W.J. & Teräsvirta T. (1993) *Modelling nonlinear economic relationships*, Oxford University Press, Oxford.

Gupta, Sunil, Lee G. Cooper. 1992. The discounting of discounts and promotion thresholds. *J. Consumer Res.* 19(Dec) 401-411.

Gustavsen, G. W., and Rickertsen K. (2006). "A Censored Quantile Regression Analysis of Vegetable Demand: The Effects of Changes in Prices and Total Expenditure." *Canadian Journal of Agricultural Economics* 54:631-45.

Huth P. J., DiRienzo D. B. & Miller G. D. (2006). Major Scientific Advances with Dairy Foods in Nutrition and Health. *J. Dairy Sci.* 89:1207-1221.

Jensen J.D. & Smed S. (2007) Cost-effective design of economic instruments in nutrition policy,

***International Journal of Behavioural Nutrition and Physical Activity* 2007, 4:10 doi:10.1186/1479-5868-4-10.**

Kahneman, Daniel. 1991. Judgment and decision making: A personal view. *Psych. Sci.* 2(3) 142-145.

Mobley, Mary F., William O. Bearden, and Jesse E. Teel (1988), "An Investigation of Individual Responses to Tensile Price Claims," *Journal of Consumer Research*, 15 (September), 273-79.

Naik N. Y. & Moore M. J. (1996). Habit formation and intertemporal substitution in individual food consumption. *The review of economics and statistics*, Vol. 78 (2), May 1996, 321-328.

Nordström, J. and L. Thunström. (2009). The Impact of tax reforms designed to encourage healthier grain consumption. *Journal of Health Economics*, 28: 622-634.

Omori Y. & Miyawaki K. (2009) Tobit model with covariate dependent thresholds, *Computational Statistics and Data Analysis* (in press).

Pauwels K., Srinivasan S. & Franses P.H. (2006). When do price thresholds matter in retail categories, *Marketing Science*.

Pereira M. A. et al. (2002). Dairy consumption, obesity, and the insulin resistance syndrome in young adults. The CARDIA study. *The journal of the American medical association*, 2002; 287(16):2081-2089.  
Doi:10.1001/jama287.16.2081.

Rosett R.N. (1959) A statistical model of friction economics, *Econometrica*, vol. 27(2), pp.263-267.

Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, In: *Econometric Society Monographs*, vol. 3, Cambridge University Press, Cambridge, New York.

Renzaho A. M. N. (2004). Fat, rich and beautiful: changing socio-cultural paradigms associated with obesity risk, nutritional status and refugee children from sub-Saharan Africa. [Health & Place](#), [Volume 10, Issue 1](#), March 2004, Pages 105-113.

Van Heerde H.J. Leeflang P.S., Wittink D.R. (2001) Semiparametric analysis to estimate the deal effect curve, *Journal of Marketing Research*, vol. 38, pp. 197-215.

Warde A.(1999) "Convenience food: space and timing," *British Food Journal*, Vol. 101 Iss: 7, pp.518 – 527.

WHO (2006). Fact sheet N°311, September 2006. Available at:

<http://www.who.int/mediacentre/factsheets/fs311/en/index.html>. Accessed on 15 Marts 2010.

Wisniewski, Kenneth J. and Robert C. Blattberg (1983), "Response Function Estimation Using UPC Scanner Data," in *Proceedings of ORSA-TIMS Marketing Science Conference*, Fred S. Zufryden, ed. Providence, RI: The Institute of Management Sciences, 300-11.

Appendix. Detailed econometric estimation results, households with or without children

	Skimmed milk	Mini milk	Semi-skimmed milk	Whole milk
<b>Children</b>				
Alpha	-859,3 (140,0)	-282,9 (263,4)	1152,8 (86,1367)	-38,9083 (49,4207)
Delta	0,1794 (0,0193)	0,0306 (0,0409)	-0,1878 (0,00857)	0,00799 (0,0195)
Gamma1	-0,7953 (0,0114)	-0,7253 (0,0109)	-0,7804 (0,00887)	-0,7808 (0,0114)
Gamma2				
Beta1	-5041,6 (497,9)	-7567,5 (596,5)	1254,2 (884,6)	-3048,8 (344,4)
Beta2	-495,3 (1006,8)	-1194,6 (796,6)	-5969,2 (372,3)	-487,5 (334,5)
S2	6973176 (121018)	6100377 (91513)	7192892 (89182,2)	2123172 (36282)
<b>No children</b>				
Alpha	-21,8359 (9,8980)	-72,3032 (31,6966)	-31,7703 (10,2647)	17,7071 (8,8306)
Delta	-0,0195 (0,00146)	0,00339 (0,00569)	-0,0116 (0,00218)	-0,0600 (0,00944)
Gamma1	-0,7216 (0,00426)	-0,7541 (0,00406)	-0,8472 (0,00336)	-0,8336 (0,00490)
Gamma2				
Beta1	-6470,9 (249,9)	-5371,7 (184,6)	-4546,9 (171,9)	-4298,8 (308,8)
Beta2	-785,6 (205,9)	198,0 (336,1)	-1113,5 (161,6)	-680,6 (131,4)
S2	2622936 (16570)	2413428 (14052)	3276162 (16075)	1501379 (10499,5)

Note: Standard deviations in parentheses